

# LARGE DEVIATIONS AND EFFECTIVE EQUIDISTRIBUTION

ILYA KHAYUTIN

**ABSTRACT.** We study large deviations for measurable averaging operators on state spaces of dynamical systems. Our main motivation is the Hecke operators on the modular curve  $Y_0(p^n)$  and their generalization to higher rank  $S$ -arithmetic quotients.

We prove a relatively sharp large deviations result in terms of the norm of the averaging operator restricted to the orthogonal complement of the constant functions in  $L^2$ . In the self-adjoint case this norm is expressible by the spectral gap.

Developing ideas of Linnik and Ellenberg, Michel and Venkatesh, we use this large deviation result to prove an effective equidistribution theorem on a state space. The novelty of our results is that they apply to measures with sub-optimal bounds on the mass of Bowen balls.

We present two new applications to our effective equidistribution result. The first one is effective rigidity for the measure of maximal entropy on  $S$ -arithmetic quotients with respect to a semisimple action in a non-archimedean place. Measures having large enough metric entropy must also be close on the state space to the Haar measure. This is a partial extension of a recent result of Rühr to a significantly more general setting.

The second one is non-escape of mass for sequences of measures having large entropy with respect to a semisimple element in a non-archimedean place. This generalizes similar known results for real flows. Our methods differ from the methods used by Rühr and in the previously known non-escape of mass results.

## CONTENTS

1. Introduction	2
2. Averaging Operators	7
3. Large Deviations	10
4. Walks on $S$ -Algebraic Quotients and Hecke Operators	15
5. Applications to Equidistribution	24
6. Effective Rigidity of the Measure of Maximal Entropy	31
7. Non-Escape of Mass	39
Appendix A. Local Entropy	43
Appendix B. Homogeneity of the Haar Measure	47
References	49

## 1. INTRODUCTION

**1.1. State Spaces and Averaging Operators.** In this paper we deal with state spaces for measurable dynamical system. A dynamical system for us is a measure space  $(Y, \mathcal{Y}, m)$  equipped with a not necessarily invertible measure-preserving transformation  $S: Y \rightarrow Y$ . A state space for the system  $Y$  is a measurable space  $(X, \mathcal{X})$  with a measurable map  $\pi: (Y, \mathcal{Y}) \rightarrow (X, \mathcal{X})$ . The space  $X$  has a natural associated measure  $\pi_* m$  – the pushforward of  $m$ ; yet the action of  $S$  does not necessarily descend to  $X$ .

One example to keep in mind is Markov shifts on finite alphabets and the associated directed graphs.

We are mainly interested in a family of dynamical systems arising in the  $S$ -arithmetic setting. Let  $\mathbf{G}$  be a reductive linear algebraic group defined over  $\mathbb{Q}$ . Let  $S$  be a finite set of places of  $\mathbb{Q}$  including  $\infty$  such that  $G_S := \prod_{v \in S} \mathbf{G}(\mathbb{Q}_v)$  is non-compact. Fix  $\Gamma < G_S$  a congruence lattice and set  $Y := \Gamma \backslash G_S$ .

Let  $p \in S$  be a *finite* place and  $a \in G_p := \mathbf{G}(\mathbb{Q}_p)$  a semisimple element not belonging to a compact group. The space  $Y$  is equipped with a unique Borel probability measure invariant under the right action of  $G_S$  – this is the Haar measure  $m$ . The dynamical system we study is  $Y$  equipped with the right action by  $a$ . A compact subgroup  $K < G_S$  defines a state space  $X := Y/K$  for this dynamical system.

Lets present an important example which we address in this paper. The space  $X := \mathbf{PGL}_2(\mathbb{Z}) \backslash \mathbf{PGL}_2(\mathbb{R})$  can be seen as a quotient by a compact group of a larger  $S$ -arithmetic locally homogeneous space, specifically

$$\begin{aligned} Y &:= \mathbf{PGL}_2(\mathbb{Z}) \backslash \mathbf{PGL}_2(\mathbb{R}) \times \mathbf{PGL}_2(\mathbb{Q}_p) \\ X &\cong Y / \mathbf{PGL}_2(\mathbb{Z}_p) \end{aligned}$$

If we consider  $Y$  as a dynamical system with the right action by  $a = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$ , then the action does not descent to the space  $X$ . Nevertheless, a ghost of this action is still visible in  $X$  – the  $p$ -Hecke operator. Our main focus is higher rank generalization of this example.

For a general dynamical system and a state space we show how the Koopman operator  $U_S$  on  $L^2(Y, m)$  induces an averaging operator  $T$  on  $L^2(Y, m)$  defined by

$$Tf := \mathbb{E}(U_S \mathbb{E}(f \mid \mathcal{X}) \mid \mathcal{X})$$

A main feature of  $T$  is that it can be restricted to an operator from  $L^2(X, m)$  to itself.

**1.2. Large Deviations.** We study state spaces for which the averaging operator  $T$  behaves in a suitable sense like the transition operator of a Markov chain. We provide the formal definition of the required property, which we call property (M), in Definition 2.6. Notice that the example of

the modular surface introduced above does not possess property (M), yet this is amended by dividing by an Iwahori subgroup of  $\mathbf{PGL}_2(\mathbb{Q}_p)$  instead of a maximal compact one. One can then deduce all the result we prove on the basic modular surface from the analogues results on this finer state space.

We develop a large deviation bound for empirical averages of functions on the state space in terms of  $\|T\|_{L_0^2(X,m)}$ , where  $L_0^2(X,m)$  is the orthogonal complement to the constant functions in  $L^2(X)$ . If the averaging operator  $T$  is self-adjoint, which is not always the case, then  $1 - \|T\|_{L_0^2(X,m)}$  is the spectral gap of  $T$ . Here is a slightly simplified version of our large deviation result.

**Theorem** (Theorem 3.2). *Let  $Y$  be a dynamical system with a state space  $X$  having property (M). Denote  $\lambda := \|T\|_{L_0^2(X,m)} \leq 1$ .*

*Let  $\varphi: Y \rightarrow [0, 1]$  be a  $\mathcal{X}$ -measurable function and  $\eta \geq m(\varphi)$ . For any  $\theta \in \{\lambda^k \mid k \in \mathbb{N}\}$  such that  $\theta \leq \frac{\eta - m(\varphi)}{1 - m(\varphi)}$*

$$m \left( y \mid \frac{1}{n} \sum_{i=0}^{n-1} \varphi(S^i y) \geq \eta \right) \leq \ll_{\theta, \lambda, \eta, m(\varphi)} \exp \left[ -n(-\log \lambda) \frac{D(\eta \parallel \theta + (1 - \theta)m(\varphi))}{-\log \theta} \right]$$

Where for any  $0 \leq a, b \leq 1$  the function  $D(a \parallel b)$  is the binary Kullback-Leibler divergence in natural base defined by

$$D(a \parallel b) := -a \log \frac{b}{a} - (1 - a) \log \frac{1 - b}{1 - a}$$

This bound applies in particular to non-reversible Markov chains in discrete time. Large deviation bounds for Markov chains with decay rate depending on spectral data of the transition operator is a well-studied subject. It has been initiated by Gillman [Gil98] and developed, inter alios, by [AFWZ95, Kah97, Lez98, LP04, Wag08, CLLM12, EMV13].

This form for the rate function of the large deviation estimate is new already in the setting of Markov chains. In particular, our large deviation theorem has the pleasant feature that for  $\lambda = 0$ , which holds for i.i.d. variables taking values in a finite alphabet, we recover the Chernoff-Hoeffding bound [Hoe63] in full strength.

**1.3. Effective Rigidity of The Measure of Maximal Entropy.** Suppose we have an  $S$ -arithmetic quotient  $\Gamma \backslash G_S$  equipped with the right action by a semisimple element in a non-archimedean place  $a \in G_p$ . As long as the group is generated by the horospherical subgroups corresponding to  $a$  – the Haar measure is known to be the unique measure of maximal entropy [MT94, §9], [EL10, §7].

We have the following form effective rigidity for the measure of maximal entropy.

**Theorem** (Theorem 6.10). *Let  $\mathbf{G}$  be a connected simply connected absolutely almost simple linear algebraic group defined over  $\mathbb{Q}$ . Let  $S$  be a finite set of places for  $\mathbb{Q}$  including  $\infty$ , such that  $G_S := \prod_{v \in S} \mathbf{G}(\mathbb{Q}_v)$  is not compact.*

*Let  $p \in S$  be a finite place such that  $G_p := \mathbf{G}(\mathbb{Q}_p)$  is not compact. Fix  $a \in G_p \cap \mathbf{G}(\mathbb{Q})$   $\mathbb{Q}_p$ -regular and semisimple and not contained in a compact subgroup.*

*Denote by  $A < G_p$  the maximal split subtorus of a maximal torus including the element  $a$ . Let  $\Delta$  be the affine Bruhat-Tits building associated to  $G_p$ . Denote by  $H < G_p$  the pointwise stabilizer of the apartment corresponding to  $A$  in  $\Delta$ .*

*Let  $\Gamma < G_S$  be a congruence lattice and denote by  $m$  the Haar probability measure on  $\Gamma \backslash G_S$ .*

*Let  $\varphi: \Gamma \backslash G_S / H \rightarrow [0, 1]$  be a continuous function that is smooth in the  $p$ -adic coordinate, i.e. when considered as a function on  $\Gamma \backslash G_S$  it is invariant under some compact open subgroup  $K_0 < G_p$  containing  $H$ .*

*Then for any  $\varepsilon > 0$  there is an explicitly computable constant  $C > 0$  depending only on  $\varepsilon$ ,  $G_p$ ,  $K_0$  and  $a \in G_p$  such that for any  $a$ -invariant measure  $\mu$  we have*

$$|\mu(\varphi) - m(\varphi)| \leq C (h_m(a) - h_\mu(a))^{1/2-\varepsilon}$$

This should be compared with the recent theorem of Rühr [Rüh15, Theorem 1.1]. The result of Rühr applies to the special case when<sup>1</sup>  $G_S = \mathbf{G}(\mathbb{Q}_p)$ ,  $\mathbf{G}$  is  $\mathbb{Q}_p$ -split and  $a$  is a semisimple element that does not belong to a compact subgroup. Wherever it applies [Rüh15, Theorem 1.1] is somewhat stronger than Theorem 6.10 and we discuss the major differences in the following.

Analogues results to Rühr's and ours are known also in other settings [Pol11, Kad14].

We note that Theorem 6.10 is derived from stronger results of more technical nature<sup>2</sup>. In particular, the methods which lead to the proof of Theorem 6.10 allow us to prove the non-escape of mass results described in the next section. Both Theorem 6.10 and a possible extension of [Rüh15, Theorem 1.1] to general  $S$ -arithmetic quotients would imply qualitatively weaker results for non-escape of mass.

Theorem 6.10 and [Rüh15, Theorem 1.1] both rely eventually on exponential decay of matrix coefficients. Bounds on matrix coefficients enter our argument because it implies a bound on the  $L_0^2$ -norm of the averaging operator [Sar91, Chi95, BS91, CU04, COU01, GMO08]. Rühr uses decay of matrix coefficients to show a uniform effective equidistribution under the action of the stable horospherical subgroup corresponding to  $a$ ; equidistribution under the horospherical action is related to mixing of the

<sup>1</sup>Necessarily  $\Gamma \backslash G_S$  is a compact quotient.

<sup>2</sup>If the measure  $\mu$  is ergodic we have a relatively concise form of the strong version of our results in Corollary 6.5.

$a$ -action as has been already observed by Margulis [Mar04]. The pertinent bounds on matrix coefficients are the result of the combined work of [BS91, CU04, Clo03, GJ78, JL70, Rog90, Oh02, Oh98].

The primary strength of our result is that it covers significantly more cases. Most importantly, we are able to treat  $S$ -arithmetic quotients with an archimedean part.

The central shortcoming of our result is that it applies only to functions invariant under  $H$ . The theorem of Rühr applies to all smooth functions. The reason we are restricted to  $H$  invariant functions is quite natural. To deduce Theorem 6.10 we need a state space  $X = Y/K$  such that  $\varphi$  factors through  $X$ . The state spaces come from subgroups  $K < G_p$  that stabilize two chambers<sup>3</sup> in the apartment corresponding to  $A$  in  $\Delta$ , see §4.3. In particular, any state space that we analyze must be coarser than the quotient by the pointwise stabilizer of this apartment –  $H$ .

Another difference is that our result applies to functions in  $L^\infty$ . If we take  $\varphi$  bounded but such that we do not have necessarily  $0 \leq \varphi \leq 1$  then  $|\mu(\varphi) - m(\varphi)|$  will be proportionate to  $\|\varphi\|_\infty$ . The result of [Rüh15] is for any smooth function in  $L^2$  and more importantly the bound is proportionate to  $\|\varphi\|_2$ . The reason our result depends on the  $L^\infty$  norm lies in the basis of the development of the large deviations bound in §3.3.

Last, we lose a factor of  $\varepsilon$  in the exponent compared to [Rüh15, Theorem 1.1]. We note that we derive Theorem 6.10 from stronger results, specifically Proposition 6.1.

**1.4. Non-Escape of Mass.** The relation between measures of high entropy and non-escape of mass has been introduced by Einsiedler, Lindenstrauss, Michel and Venkatesh [ELMV12] and extended in [EK12], [EKP11], [Kad12a] and [KKLM14]. Closely related is the work of Cheung [Che11]. These results described non-escape of mass for flows under a one-parameter subgroup in the real place<sup>4</sup>.

We present a theorem that is closely related to the previously studied cases. Our result establishes non-escape of mass for measures of high entropy invariant under the action of a regular semisimple element in the  $p$ -adic place. The tools we use in proving this theorem are considerably different from the methods applied in the proofs of the previously established results. Specifically, none of the above relied on decay of matrix coefficients.

Assume that  $a \in G_p$  is  $\mathbb{Q}_p$ -regular and semisimple. Let  $A < G_p$  be the maximal split subtorus of a maximal torus containing  $a$  and denote by  $A^+$  the closed Weyl chamber corresponding to  $a$ . The function  $\eta: A \rightarrow \mathbb{R}$  has been defined by Oh [Oh98] and extended in [COU01] and [GMO08]. It is equal to the product of all the roots in a maximal strongly orthogonal system

<sup>3</sup>Not necessarily distinct.

<sup>4</sup>Actions in rank 1 in a  $p$ -adic place have been briefly discussed in [ELMV12, Remark 5.2].

of positive roots, see Definition 7.3. This function is closely related to the decay of matrix coefficients and the norm gap of Hecke operators.

For the diagonal torus in  $\mathbf{SL}_n$  and

$$A^+ = \{\text{diag}(a_1, \dots, a_n) \mid a_i \in \mathbb{N}, a_i > 0 \text{ and } \forall 1 \leq i \leq n-1: a_{i+1} \mid a_i\}$$

the function  $\eta$  for  $a \in A^+$  is equal to

$$\eta(a) = \prod_{i=1}^{\lfloor n/2 \rfloor} \frac{|a_i|_p}{|a_{n+1-i}|_p}$$

**Theorem** (Theorem 7.5). *Let  $\mathbf{G}$  be a connected simply connected absolutely almost simple linear algebraic group defined over  $\mathbb{Q}$ . Let  $S$  be a finite set of places for  $\mathbb{Q}$  including  $\infty$ , such that  $G_S := \prod_{v \in S} \mathbf{G}(\mathbb{Q}_v)$  is not compact.*

*Let  $p \in S$  be a finite place such that  $G_p := \mathbf{G}(\mathbb{Q}_p)$  is not compact, set  $r = 2$  if  $\text{rank}_{\mathbb{Q}_p} \mathbf{G} = 1$  and  $r = 1$  if the rank is higher.*

*Fix  $a \in G_p \cap \mathbf{G}(\mathbb{Q})$  a  $\mathbb{Q}_p$ -regular and semisimple element not belonging to a compact subgroup. Let  $A < G_p$  be the maximal split subtorus of a maximal torus containing  $a$ , set  $A^+$  to be the closed Weyl chamber corresponding to  $a$ . Define  $\eta$  with respect to these choices.*

*Let  $\Gamma < G_S$  be a congruence lattice and denote by  $m$  the Haar probability measure on  $\Gamma \backslash G_S$ .*

*Suppose we are given a sequence of  $a$ -invariant probability measures  $\mu_i$  on  $Y := \Gamma \backslash G$ . Set  $h = \liminf_{i \rightarrow \infty} h_{\mu_i}(a)$ . If  $\mu_i$  converges to a measure  $\mu$  in the weak-\* topology then*

$$\mu(Y) \geq 1 - 2r \frac{h_m(a) - h}{-\log \eta(a)}$$

The factor  $r = 2$  in rank 1 is related to the Ramanujan conjecture [CU04] and is conjectured to be redundant [GMO08, Conjecture 2.15].

Although our results are for the  $p$ -adic place our expression for the bound on the mass is well defined in the real place as well. It is interesting to compare this bound with the established bounds in the real place and especially with [Kad12b], [KP15] which establish that some of the known bounds for real flows are sharp.

In the known archimedean cases with rank  $> 1$  the expression  $1 - 2 \frac{h_m(a) - h}{-\log \eta(a)}$  is of the right order of magnitude but weaker then the known results.

Lets consider the the diagonal flow

$$a(t) = \text{diag}(e^t, e^t, \dots, e^{-nt})$$

on  $\mathbf{SL}_{n+1}(\mathbb{Z}) \backslash \mathbf{SL}_{n+1}(\mathbb{R})$ . It is established in [KKLM14] that if sequence of  $a(t)$  invariant measures  $\mu_i$  converges to  $\mu$  then  $\mu(Y) \geq h/n - n$  where  $h = \liminf_{i \rightarrow \infty} h_{\mu_i}(a(1))$ . Moreover, Kadyrov [Kad12b] shows that this bound is actually sharp. The value of  $1 - 2r \frac{h_m(a) - h}{-\log \eta(a)}$  in this case is  $2 \left( \frac{h}{n+1} - n \right) + 1$  which is strictly smaller then the sharp bound of [KKLM14].

**1.5. Effective Equidistribution.** The bridge between large deviation estimates and our results for  $S$ -arithmetic quotients is provided by an effective equidistribution result, Theorem 5.10. This theorem is closely related to the work of Linnik [Lin68] and Ellenberg, Michel and Venkatesh [EMV13].

The method of proof for Theorem 5.10 uses the large deviation bound to show that if a measure  $\mu$  on an  $S$ -arithmetic quotient  $\Gamma \backslash G_S$  has on average small Bowen balls for the right action by  $a$ , then on the state space  $\Gamma \backslash G_S / K$  the measure  $\mu$  must be close in a suitable sense to the Haar measure. The novelty of our result compared to [Lin68] and [EMV13] is that we are able to provide non-trivial information even for a sub-optimal bound on the average measure of Bowen balls.

In the works of Linnik and Ellenberg, Michel, and Venkatesh the Haar measure of a set of atypical points is bounded above by an exponentially decaying bound coming from a large deviation estimate. The same measure is bounded below by a sub-exponential bound derived from an asymptotically optimal estimate on the average size of Bowen balls. The inequality between an exponential and a sub-exponential expression provides a contradiction no matter how good the rate function is in the large deviations estimate, as long as we have exponential decay.

When the available estimate on the average measure of Bowen balls is sub-optimal the lower bound is no longer sub-exponential. In this case we have lower and upper bounds both of exponential type. In this case, to derive a contradiction we need to compare the rate of decay of the two bounds. For this we require that the rate of the large deviation estimate be as good as possible.

The better exponential rate we have in the pertinent large deviations bound the better equidistribution result we have using the methods of Theorem 5.10.

**1.6. Acknowledgments.** This paper is part of the author's PhD thesis conducted at the Hebrew University of Jerusalem under the guidance of Prof. E. Lindenstrauss, to whom I am grateful for introducing me to homogeneous dynamics and number theory, and for many helpful discussions and insights. I would like to express my gratitude to Akshay Venkatesh for encouraging and valuable conversations. It is my pleasure to thank Or Landesberg and Shai Evra for helpful discussions and much needed coffee breaks.

Last but not least, I would like to thank my wife, Olga Kalantarov, for her enduring support during the preparation of this paper.

The author has been supported by the European Research Council [AdG Grant 267259] throughout his PhD studies.

## 2. AVERAGING OPERATORS

In this section we present a general framework for studying contraction properties for averaging operators coming from dynamical systems.

First we describe how a measurable dynamical system and a state space gives rise to an averaging operator.

**Definition 2.1.** Let  $(Y, \mathcal{Y}, m, S)$  be a dynamical system, viz.  $(Y, \mathcal{Y}, m)$  is a measure space and  $S: Y \rightarrow Y$  is a measure preserving transformation. We define a *state space* for the system  $Y$  to be a measurable space  $(X, \mathcal{X})$  with a measurable map  $\pi: Y \rightarrow X$ . The state space has a natural measure on it  $\pi_* m$ .

We identify  $\mathcal{X}$  with the  $\sigma$ -algebra  $\pi^{-1}\mathcal{X} \subseteq \mathcal{Y}$ . State spaces for  $Y$  are in bijection with  $m$ -equivalence classes of sub- $\sigma$ -algebras of  $\mathcal{Y}$ . Notice that  $S$  doesn't necessarily act on  $X$ , equivalently  $\mathcal{X}$  is not necessarily  $S$ -stable.

**Definition 2.2.** Let  $U_S$  be the Koopman operator associated with  $S$  acting either on  $L^1(Y)$  or on  $L^2(Y)$ . The averaging operator  $T_S$  is defined on  $L^1(Y)$  and on  $L^2(Y)$  by  $T_S f = \mathbb{E}(U_S \mathbb{E}(f | \mathcal{X}) | \mathcal{X})$ . If  $P_X$  is the orthogonal projection onto  $L^2(X) \leq L^2(Y)$  then  $T_S$  on  $L^2(Y)$  can be written as  $T_S = P_X U_S P_X$ . In functional analysis literature the operator  $T_S$  on  $L^2(Y)$  is called the compression of  $U_S$  onto  $L^2(X)$ . In particular,  $T_S$  is a well defined operator on  $L^1(X)$  and  $L^2(X)$ .

**Example 2.3.** A finite state Markov chain with the shift map is a dynamical system. The associated directed graph is a state space for this dynamical system. The averaging operator in this case is the weighted normalised adjacency matrix.

**Example 2.4.** The space

$$Y = \mathbf{PGL}_2 \left( \mathbb{Z} \begin{bmatrix} 1 \\ p \end{bmatrix} \right) \backslash \mathbf{PGL}_2(\mathbb{R}) \times \mathbf{PGL}_2(\mathbb{Q}_p)$$

is a dynamical system when equipped with the  $p$ -adic diagonal right action by  $\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$ . The space

$$X = \mathbf{PGL}_2(\mathbb{Z}) \backslash \mathbf{PGL}_2(\mathbb{R})$$

can be identified with  $Y/K$  where  $K = \mathbf{PGL}_2(\mathbb{Z}_p)$ . In particular, there is a continuous projection  $Y \rightarrow X$  and we consider  $X$  as a state space for the dynamical system  $Y$ .

The associated averaging operator is the well-known  $p$ -Hecke operator.

**Example 2.5.** Lets consider a compact Riemannian manifold  $M$  and let  $g_t$  be the geodesic flow on  $T_1 M$ . Fix  $t \in \mathbb{R}$ , the dynamical system we are interested in is the  $t$ -time map of the geodesic flow on  $T_1 M$  with the Liouville measure. The state space is  $M$  and  $T_{g_t}$  is closely related to the Laplacian which is up to a constant equal to  $\lim_{t \rightarrow 0} \frac{1}{t^2}(T_{g_t} - Id)$ .

Those examples demonstrate that although the Koopman operator being a unitary operator is as far as an operator can be from having a spectral gap, its compression to an averaging operator on a state space can have a



spectral gap. A necessary but not a sufficient condition for this is that the state space's  $\sigma$ -algebra has no  $S$ -stable sub- $\sigma$ -algebra.

More specifically, we are going to discuss not the spectral gap of the operator  $T_S$  but rather its  $L_0^2$ -norm. The space  $L_0^2(Y, m)$  is the orthogonal complement to the constant function in  $L^2(Y, m)$ . If  $T_S$  is self-adjoint then the quantity  $1 - \|T_S\|_{L_0^2(Y, m)}$  is the spectral gap of  $T_S$ .

**2.1. Property (M).** When the dynamical system  $Y$  is a 1-step Markov process with the associated directed graph as a state space, it holds for all  $n \in \mathbb{N}$  that  $T_{S^n} = (T_S)^n$ , i.e. the operators  $\{T_{S^n}\}_{n \in \mathbb{N}}$  form a semigroup.

It is important to note that it is not always true that the averaging operators form a semigroup. They do not form a semigroup<sup>5</sup> in Example 2.4, but they almost do so. We need to divide on the right not by a maximal compact subgroup of  $\mathbf{PGL}_2(\mathbb{Q}_p)$  but by an Iwahori subgroup. The associated state space is  $\Gamma \backslash \mathbf{PGL}_2(\mathbb{R})$  where  $\Gamma$  is a congruence lattice defined by  $\Gamma = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbf{PGL}_2(\mathbb{Z}) \mid c \equiv 0 \pmod{p} \right\}$ .

The primary property we use for the derivation of a large deviations bound is described in the following definition. This property is possessed by 1-step Markov processes.

**Definition 2.6.** Let  $(Y, \mathcal{Y}, m, S)$  be a dynamical system. Let  $(X, \mathcal{X})$  be a state space for  $Y$ . We say that the system  $Y$  has *property (M)* with respect to the state space  $X$  if there exists a closed  $U_S$ -stable  $L^\infty(X)$ -module<sup>6</sup>  $\mathcal{M} \subseteq L^2(Y)$  such that  $L^2(X) \subseteq \mathcal{M}$  and

$$(P_X U_S P_X)|_{\mathcal{M}} = (P_X U_S)|_{\mathcal{M}}$$

Equivalently, the orthogonal complement of  $L^2(X)$  in  $\mathcal{M}$  is  $U_S$ -stable.

*Remark 2.7.* Property (M) for  $S$  implies the same for  $S^k$  for any  $k \in \mathbb{N}$ .

*Remark 2.8.* Assume the transformation  $S$  is invertible, i.e. we have a dynamical system with a  $\mathbb{Z}$ -action.

In the interesting case that the state space is not invariant under the action of  $S$ , if property (M) holds then on  $\mathcal{M}$  the Koopman operator  $U_S$  would be non-invertible, while on  $L^2(Y)$  as a whole it is invertible. If  $U_S$  were invertible over  $\mathcal{M}$ , then the  $U_S$ -stability of the orthogonal complement of  $L^2(X)$  in  $\mathcal{M}$  would imply the  $U_S$ -stability of  $L^2(X)$  which is equivalent to  $X$  being  $S$ -invariant.

Dynamical systems with property (M) over a state space have the following important properties.

**Proposition 2.9.** *Let  $(Y, \mathcal{Y}, m, S)$  be a dynamical system and  $(X, \mathcal{X})$  a state space for  $Y$  with property (M).*

<sup>5</sup>It is well known for  $p$ -Hecke operators on the modular surface that  $T_{p^2} \neq (T_p)^2$

<sup>6</sup>A closed  $L^\infty(X)$ -module in  $L^2(Y)$  is a closed subspace of  $L^2(Y)$  stable under multiplication by any member of  $L^\infty(X)$ .

- (1)  $T_S^n = (T_S)^n$  for all  $n \geq 1$   
(2) For  $h \in L^\infty(X)$  define  $M_h$  to be the operator over  $L^2(Y)$  of multiplying by  $h$ . Then  $P_X(M_h U_S M_h)^n P_X = (M_h T_S M_h)^n$  for all  $n \geq 1$

*Proof.* The second claim is a generalization of the first so it is enough to prove it.

Let  $\mathcal{M} \leq L^2(Y)$  be as in the definition of property (M). For each  $f \in L^2(Y)$  the vector  $M_h(M_h U_S M_h)^{n-1}(P_X f) \in \mathcal{M}$ . In addition,  $P_X M_h = M_h P_X$ . By induction on <sup>7</sup> $n$

$$\begin{aligned}
P_X(M_h U_S M_h)^n P_X f &= M_h P_X U_S \left( M_h (M_h U_S M_h)^{n-1} P_X f \right) \\
&= M_h P_X U_S P_X M_h \left( (M_h U_S M_h)^{n-1} P_X f \right) \\
&= M_h (P_X U_S P_X) P_X M_h \left( (M_h U_S M_h)^{n-1} P_X f \right) \\
&= M_h T_S P_X M_h \left( (M_h U_S M_h)^{n-1} P_X f \right) \\
&= M_h T_S M_h P_X \left( (M_h U_S M_h)^{n-1} P_X f \right) \\
&= (M_h T_S M_h)^n f
\end{aligned}$$

□

### 3. LARGE DEVIATIONS

**3.1. Standing Assumptions.** Let  $(Y, \mathcal{Y}, m, S)$  be a dynamical system with a state space  $(X, \mathcal{X})$  having property (M). We denote the associated averaging operator by  $T := T_S$ .

Let  $\varphi: Y \rightarrow \mathbb{R}_+$  be a bounded  $\mathcal{X}$ -measurable function. Set  $A_n(y) := \frac{1}{n} \sum_{i=0}^{n-1} \varphi(S^i y)$ .

**3.2. Exponential Moments and the Logarithmic Generating Function.** We would like to bound  $m(y \in Y \mid A_n(y) \geq \eta)$ . We follow the standard scheme of proving large deviation bounds, like Cramér's theorem, using a bound on the logarithmic generating function, see [DZ10, §§2.2-2.3], [Kah97, Proof of Lemma 3.2]. Notice that we do not assume  $T$  to be self-adjoint.

The approach we take is to bound the exponential moment of the empiric average over the times  $0, \dots, n-1$  by a logarithmic generating function. An alternative approach to obtaining a large deviations bound is to use standard geometric moments of degree  $n$  for arbitrary large  $n$ . This is the approach taken in [AFWZ95, §4] and [EMV13, §12], see also [HLW06, §3.3].

---

<sup>7</sup>The case  $n=1$  is similar to the following.

The Markov inequality with the exponential function implies the following for any real  $r > 0$ .

$$\begin{aligned}
 m(y \in Y \mid A_n(y) \geq \eta) &= m(y \in Y \mid \exp(rnA_n(y)) \geq \exp(rn\eta)) \\
 &\leq \int \exp(rnA_n) dm \exp(-rn\eta) \\
 (1) \quad &= \exp \left\{ -n \left( r\eta - \frac{1}{n} \log \int \exp(rnA_n) dm \right) \right\}
 \end{aligned}$$

We define the logarithmic generating function and its convex conjugate

$$\begin{aligned}
 \Lambda_n(r) &= \log \int \exp(rA_n) dm \\
 \Lambda_n^*(\eta) &= \sup_{r \in \mathbb{R}} \left( r\eta - \frac{1}{n} \Lambda_n(rn) \right)
 \end{aligned}$$

the asymptotic exponential rate in the former bound is given by

$$\begin{aligned}
 \Lambda(r) &= \limsup_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(nr) \\
 \Lambda^*(\eta) &= \sup_{r \in \mathbb{R}} (r\eta - \Lambda(r))
 \end{aligned}$$

If we bound  $\Lambda_n$  we can bound not only  $\Lambda^*$  but actually bound effectively the large deviations probability for a given  $n$ . For this we use property (M) and the fact that  $\varphi$  is  $\mathcal{X}$ -measurable.

$$\begin{aligned}
 \int \exp(rnA_n) dm &= \int \prod_{i=0}^{n-1} (U^i \exp(r\varphi)) dm \\
 &= \int \exp(r\varphi) \cdot (U M_{\exp(r\varphi)})^{n-1} 1 dm \\
 &= \int \exp(r\varphi/2) \cdot (M_{\exp(r\varphi/2)} U M_{\exp(r\varphi/2)})^{n-1} \exp(r\varphi/2) dm \\
 &= \left\langle \exp(r\varphi/2), (M_{\exp(r\varphi/2)} U M_{\exp(r\varphi/2)})^{n-1} \exp(r\varphi/2) \right\rangle \\
 &= \left\langle \exp(r\varphi/2), (M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)})^{n-1} \exp(r\varphi/2) \right\rangle \\
 (2) \quad &\leq \|\exp(r\varphi/2)\|_2^2 \|M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)}\|_{L^2(X)}^{n-1}
 \end{aligned}$$

We are left with bounding  $\|M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)}\|_{L^2(X)}$  which we are going to do using a simple method, an elementary variant of which is essentially due to Kahale [Kah91, Lemma 3]<sup>8</sup>.

Notice that instead of using the norm  $\|M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)}\|_{L^2(X)}$  we could have by Gelfand's formula bound the asymptotic large deviations rate

---

<sup>8</sup>Kahale has computed a large deviations bound for the probability that a random walk on a undirected finite regular graph stays all the time in a subgraph. He has also shown that his bound is optimal in a specific sense.

by the spectral radius

$$\rho \left( M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)} \right)_{L^2(X)}$$

Actually if  $T$  is compact then as a consequence of the Krein-Rutman theorem [KR48, KR50]<sup>9</sup> the spectral radius is not only a bound but actually  $\Lambda(r) = \rho \left( M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)} \right)_{L^2(X)}$ , at least for some functions  $\varphi$ . Furthermore the Gärtner-Ellis theorem of large deviations implies under some restrictions that the function  $\Lambda^*(\eta)$  is not only an upper bound on the large deviations probability but also a lower bound in a suitable sense.

If  $T$  is not only compact but also self-adjoint then  $M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)}$  is also compact self-adjoint and its operator norm is equal to the spectral radius.

**3.3. Generalization of Kahale's Method.** A large deviation estimate without a self-adjointness assumption has been shown by Ellenberg, Michel and Venkatesh in [EMV13, Proposition 12.4]. Our motivation is mainly based upon Linnik's application of large deviation to equidistribution [Lin68] and the development on his method by Ellenberg, Michel and Venkatesh.

The bound of [EMV13] is for occupation times of subsets in a random walk on a directed finite graph. Both our method, which is more general and provides somewhat stronger results, and the method of Ellenberg, Michel and Venkatesh can be traced back to Kahale [Kah91]. The method of [EMV13] builds upon [AFWZ95] while we use a classical approach using exponential moments.

From now on we assume that  $0 \leq \varphi \leq 1$   $m$ -almost everywhere. Notice that for any  $r \geq 0$

$$(3) \quad 0 \leq e^{r\varphi} = 1 + \sum_{n=1}^{\infty} \varphi^n \frac{r^n}{n!} \leq 1 + \sum_{n=1}^{\infty} \varphi \frac{r^n}{n!} = 1 + \varphi(e^r - 1)$$

The inequalities can be replaced with equalities if  $\varphi = \chi_A$  for some  $A \in \mathcal{X}$ .

For any  $f \in L^2(X)$  because  $0 \leq \varphi \leq 1$  almost anywhere

$$(4) \quad \|\exp(r\varphi/2)f\|_2 \leq \|\exp(r/2)f\|_2 = e^{r/2}\|f\|_2$$

Assume that  $L^2(X) = V \oplus V^\perp$ , where  $V \subseteq L^2(X)$  is a closed subspace such that both  $V$  and  $V^\perp$  are  $T$ -stable<sup>10</sup>. For any  $f \in L^2(X)$  denote by  $(f)_V$  and  $(f)_{V^\perp}$  the orthogonal projections of  $f$  to  $V$  and  $V^\perp$  respectively.

We are going to bound the norm of  $M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)}$  by its norm on  $V$  and on  $V^\perp$ . Assume we know that  $\|T\|_{V^\perp} \leq \lambda$  for some  $1 > \lambda > 0$  but on  $V$  we only know  $\|T\|_V \leq 1$ .

<sup>9</sup>analogue of Perron-Frobenius for compact operators

<sup>10</sup>As we don't assume  $T$  to be self-adjoint stability of  $V$  under  $T$  does not immediately imply the same for  $V^\perp$ .

The operator norm can be calculated using

$$\|M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)}\|_{L^2(X)} = \sup_{\substack{f, g \in L^2(X) \\ \|f\|_2 = \|g\|_2 = 1}} \langle g, M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)} f \rangle$$

Let  $f, g \in L^2(X)$  such that  $\|f\|_2 = \|g\|_2 = 1$ , then

$$\begin{aligned} \langle g, M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)} f \rangle &= \langle M_{\exp(r\varphi/2)} g, T M_{\exp(r\varphi/2)} f \rangle = \\ &= \langle e^{r\varphi/2} g, T e^{r\varphi/2} f \rangle = \\ &= \langle (e^{r\varphi/2} g)_{V^\perp}, T[(e^{r\varphi/2} f)_{V^\perp}] \rangle + \langle (e^{r\varphi/2} g)_V, T[(e^{r\varphi/2} f)_V] \rangle \leq \\ &\leq \lambda \|(e^{r\varphi/2} g)_{V^\perp}\|_2 \|(e^{r\varphi/2} f)_{V^\perp}\|_2 + \|(e^{r\varphi/2} g)_V\|_2 \|(e^{r\varphi/2} f)_V\|_2 \leq \\ &\leq \lambda \frac{\|(e^{r\varphi/2} g)_{V^\perp}\|_2^2 + \|(e^{r\varphi/2} f)_{V^\perp}\|_2^2}{2} + \|(e^{r\varphi/2} g)_V\|_2 \|(e^{r\varphi/2} f)_V\|_2 \end{aligned}$$

The last line follows from the inequality of the means. We can now write  $\|(e^{r\varphi/2} f)_{V^\perp}\|_2^2 = \|e^{r\varphi/2} f\|_2^2 - \|(e^{r\varphi/2} f)_V\|_2^2$  and the same with  $g$ . Using this substitution and applying again the inequality of the means and (4) we have

$$\begin{aligned} &\langle g, M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)} f \rangle \\ &\leq \lambda \frac{\|e^{r\varphi/2} g\|_2^2 + \|e^{r\varphi/2} f\|_2^2}{2} + (1 - \lambda) \|(e^{r\varphi/2} g)_V\|_2 \|(e^{r\varphi/2} f)_V\|_2 \\ (5) \quad &\leq \lambda e^r + (1 - \lambda) \|(e^{r\varphi/2} g)_V\|_2 \|(e^{r\varphi/2} f)_V\|_2 \end{aligned}$$

**3.4.  $L_0^2$ -Norm.** Take  $V = 1\mathbb{C}$  to be the subspace of constant functions. If  $T$  is self-adjoint then  $1 - \lambda$  is the spectral gap of  $T$ . Furthermore, for any  $f \in L^2(X)$  we have by Cauchy-Schwartz

$$(6) \quad \|(e^{r\varphi/2} f)_V\|_2 = \left| \int e^{r\varphi/2} f \, dm \right| \leq \|e^{r\varphi/2}\|_2 \|f\|_2 = \|e^{r\varphi/2}\|_2$$

Let  $m(\varphi) = \int \varphi \, dm$ , then by (6) and (3)

$$\|(e^{r\varphi/2} g)_V\|_2 \|(e^{r\varphi/2} f)_V\|_2 \leq \|e^{r\varphi/2}\|_2^2 = \int e^{r\varphi} \, dm \leq 1 + m(\varphi)(e^r - 1)$$

The final bound on the norm implied by (5) and the computation above is

$$\|M_{\exp(r\varphi/2)} T M_{\exp(r\varphi/2)}\|_{L^2(X)} \leq 1 + [\lambda + (1 - \lambda)m(\varphi)](e^r - 1)$$

Let  $\beta = \lambda + (1 - \lambda)m(\varphi)$ . Notice that  $m(\varphi) \leq \beta$  because  $0 \leq m(\varphi) \leq 1$ . Combining  $\|\exp(r\varphi/2)\|_2^2 \leq 1 + m(\varphi)(e^r - 1)$  and our norm bound with (3) we have

$$\int \exp(rn A_n) \, dm \leq [1 + m(\varphi)(e^r - 1)][1 + \beta(e^r - 1)]^{n-1} \leq [1 + \beta(e^r - 1)]^n$$

By optimizing  $r$  we have for  $\eta \geq \beta$  that

$$\Lambda_n^*(\eta) \geq D(\eta \| \beta)$$

$$D(\eta \| \beta) := -\eta \log \frac{\beta}{\eta} - (1 - \eta) \log \frac{1 - \beta}{1 - \eta}$$

For all  $a, b \in [0, 1]$  the function  $D(a \| b)$  is the Kullback-Leibler divergence of two probability distributions on the space of two points, the distributions being  $(a, 1 - a)$  and  $(b, 1 - b)$ .

Combining the last bound with (1) and (2) we see that for  $\eta \geq \beta$

$$(7) \quad m(y \in Y \mid A_n(y) \geq \eta) \leq \exp[-nD(\eta \| \beta)]$$

If  $\varphi = \chi_A$  for some  $A \in \mathcal{X}$  then this bound is exactly the Chernoff-Hoeffding large deviations theorem for i.i.d. random variables for occupation of a set of measure  $\beta$  instead of  $m(A)$ . A pleasant feature of our bound is that by taking  $\lambda = 0$  which is the case for i.i.d. random variables<sup>11</sup> we recover the Chernoff-Hoeffding theorem in full strength.

**3.4.1. Norm Exponentiation.** What happens if  $\eta > m(\varphi)$  but  $\eta \leq \beta := \lambda + (1 - \lambda)m(\varphi)$ ? In this case the results of the previous section do not apply verbatim. We follow an established strategy to overcome this difficulty which consists considering the  $S^k$  dynamics for high enough  $k$ , [EMV13, Proof of Lemma 12.1.13].

The bound from the previous section can be improved for  $\varphi$  with  $m(\varphi)$  small by looking at the  $S^k$  dynamics for a fixed  $k$ . Remember that property (M) implies  $T_{S^k} = (T_S)^k$ , hence it has a smaller  $L_0^2$  norm, i.e.  $\|T^k\|_{V^\perp} \leq \lambda^k$ .

For  $y \in Y$  for which  $A_n(y) \geq \eta$  we can for any  $1 \leq k \leq n$  split the time series  $0, 1, \dots, n - 1$  into  $k$  arithmetic progressions of length  $> n/k - 1$  such that for at least one of them the empirical average of  $\phi$  is  $\geq \eta$ . By applying estimate (7) to each such arithmetic progression we have for any<sup>12</sup>  $k \in \mathbb{N}$  such that  $\eta \geq \lambda^k + (1 - \lambda^k)m(\varphi)$

$$(8) \quad m(y \in Y \mid A_n(y) \geq \eta) \leq k \exp \left[ -n \left( \frac{1}{k} - \frac{1}{n} \right) D(\eta \| \lambda^k + (1 - \lambda^k)m(\varphi)) \right]$$

$$= k \exp \left[ D(\eta \| \lambda^k + (1 - \lambda^k)m(\varphi)) \right] \exp \left[ -n \frac{D(\eta \| \lambda^k + (1 - \lambda^k)m(\varphi))}{k} \right]$$

$$\leq k \exp [D(\eta \| m(\varphi))] \exp \left[ -n \frac{D(\eta \| \lambda^k + (1 - \lambda^k)m(\varphi))}{k} \right]$$

In the last line we have used that if  $\eta \geq \lambda^k + (1 - \lambda^k)m(\varphi) \geq m(\varphi)$  then  $D(\eta \| \lambda^k + (1 - \lambda^k)m(\varphi)) \leq D(\eta \| m(\varphi))$ .

<sup>11</sup>The corresponding dynamical system is just the one-sided Bernoulli shift and the state space is the the alphabet of the shift. The map between the two is the projection of a sequence to its first entry.

<sup>12</sup>If  $k > n$  the following bound is trivially correct as in this case  $n/k - 1 < 0$ .

*Remark 3.1.* The multiplicative factor  $\exp D(\eta \| m(\varphi))$  can be optimized a little bit further by noticing that the length of the arithmetic progression is actually  $\geq n/k - (k-1)/k$  and keeping the dependence on  $k$ . As our main interest lies in the exponential rate constant for large  $n$ , this improvement will be negligible for us and will be unnecessary notationally cumbersome.

We now make a comfortable change of variables. We denote  $\theta = \lambda^k$ , then  $k = \log \theta / \log \lambda$  and the condition  $k \in \mathbb{N}$  becomes  $\theta \in \{\lambda^k \mid k \in \mathbb{N}\}$ . For convenience's sake we use the non-standard notation  $\lambda^{\mathbb{N}} := \{\lambda^k \mid k \in \mathbb{N}\}$ . Notice that  $\lambda^{\mathbb{N}} \subseteq (0, \lambda]$ .

**Theorem 3.2.** *Let  $\phi: Y \rightarrow [0, 1]$  be a  $\mathcal{X}$ -measurable function. Let  $T$  be the averaging operator corresponding to  $U_S$  on  $X$ . Denote  $\lambda := \|T\|_{L_0^2(X)} \leq 1$ .*

*For any  $\theta \in \lambda^{\mathbb{N}} := \{\lambda^k \mid k \in \mathbb{N}\} \subseteq (0, \lambda]$  we have*

$$\begin{aligned}
 & \text{(a) If } \eta \geq m(\varphi) \text{ and } \theta \leq \frac{\eta - m(\varphi)}{1 - m(\varphi)} \text{ then} \\
 & m \left( y \mid \frac{1}{n} \sum_{i=0}^{n-1} \varphi(S^i y) \geq \eta \right) \leq \frac{\log \theta}{\log \lambda} \exp D(\eta \| m(\varphi)) \\
 & \quad \times \exp \left[ -n(-\log \lambda) \frac{D(\eta \| \theta + (1 - \theta)m(\varphi))}{-\log \theta} \right] \\
 & \text{(b) If } \eta \leq m(\varphi) \text{ and } \theta \leq \frac{m(\varphi) - \eta}{m(\varphi)} \text{ then} \\
 & m \left( y \mid \frac{1}{n} \sum_{i=0}^{n-1} \varphi(S^i y) \leq \eta \right) \leq \frac{\log \theta}{\log \lambda} \exp D(\eta \| m(\varphi)) \\
 & \quad \times \exp \left[ -n(-\log \lambda) \frac{D(\eta \| (1 - \theta)m(\varphi))}{-\log \theta} \right]
 \end{aligned}$$

*Proof.* Part (a) is (8). Part (b) is part (a) applied to the function  $1 - \varphi$ .  $\square$

#### 4. WALKS ON $S$ -ALGEBRAIC QUOTIENTS AND HECKE OPERATORS

In this section we show that some walks on  $S$ -algebraic quotients fall under the framework of dynamical systems with a state space having property (M). Moreover, the associated averaging operators will be variants of the classical Hecke operators on the modular surface.

The archetypal example is the walks on the  $p$ -Hecke graph embedded in

$$X := \mathbf{PGL}_2(\mathbb{Z}) \backslash \mathbf{PGL}_2(\mathbb{R})$$

Set

$$Y := \mathbf{PGL}_2(\mathbb{Z}[1/p]) \backslash \mathbf{PGL}_2(\mathbb{R}) \times \mathbf{PGL}_2(\mathbb{Q}_p)$$

and equip  $Y$  with the right action of the diagonal element  $a = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \in \mathbf{PGL}_2(\mathbb{Q}_p)$ . Denote  $K := \mathbf{PGL}_2(\mathbb{Z}_p)$ , we recall that there is a natural identification  $X \cong Y/K$ . Let  $\pi: Y \rightarrow X$  be the natural projection  $y \mapsto yK$ .

One can understand the Hecke neighbors of a point  $x \in X$  by taking its lift to  $Y$  which is the set  $xK$ , apply  $a$  to have  $xKa^{-1}$  and project back to  $X$  which results in  $\pi(xKa^{-1}) = xKa^{-1}K$ . The double coset  $Ka^{-1}K$  can be written as finite union of size  $p + 1$  of right  $K$  cosets. This gives rise to exactly  $p + 1$  Hecke neighbors of  $x \in X$ .

More generally, each point  $y \in Y$  projects to a walk on the Hecke graph in  $X$  emanating from  $\pi(x) \in X$ . this walk is given by

$$(\dots, \pi(ya^2), \pi(ya), \pi(y), \pi(ya^{-1}), \pi(ya^{-2}), \dots)$$

One would like to use the spectral gap results for the  $p$ -Hecke operator to derive a large deviation result in this setting. The flavor of the results we are after is that the measure of points  $y \in Y$  whose walks of length  $n$  are all contained in a small ball in  $X$  decays exponentially with  $n$ .

Unfortunately, the state space  $X$  does not have property (M) for the system  $Y$  with the right  $a$ -action. In particular, the results of Proposition 2.9 do not hold for the corresponding averaging operator – the  $p$ -Hecke operator. For example,  $T_{p^2} \neq (T_p)^2$ .

To amend the situation one replaces  $X$  with a finer state space. This is achieved by dividing  $Y$  on the right by a smaller compact open subgroup – the Iwahori subgroup  $I$ .

More generally one can consider walks on quotients of  $S$ -algebraic groups associated to Bruhat-Tits buildings, we construct such walks explicitly and study their relevant properties in §4.3.

#### 4.1. General Setting.

4.1.1. *Standing Assumptions.* Let  $G$  be a second-countable locally compact topological group, such a group is necessarily metrizable with a left invariant proper<sup>13</sup> metric [Str74]. Notably, it is  $\sigma$ -compact. Fix a lattice<sup>14</sup>  $\Gamma < G$ , set  $Y := \Gamma \backslash G$ . Recall that  $\Gamma$  acts freely and properly discontinuously on  $G$ , in particular, the space  $Y$  is locally compact and a left invariant metric on  $G$  induces a metric on  $Y$ .

Denote by  $\mathcal{Y}$  the Borel  $\sigma$ -algebra of  $Y$ . The space  $Y$  carries a unique  $G$  invariant probability measure which we denote by  $m$ . The space  $Y$  is not necessarily compact. All  $L^q$  spaces will be with respect to this measure  $m$ .

Let  $K < G$  be a compact subgroup. Suppose that  $N_G(K) < G$  is an open subgroup of  $G$ . We define  $X := Y/K = \Gamma \backslash G/K$  and denote by  $\pi: Y \rightarrow X$  the continuous quotient map. Set  $\mathcal{X}$  to be the Borel  $\sigma$ -algebra of  $X$  which we treat as a subalgebra of  $\mathcal{Y}$  using the map  $\pi$ .

The measure  $m$  can be pushed forward to  $X$  using  $\pi$ ; by abuse of notation, we denote the pushed forward measure by  $m$  as well. Because of the universal property of the quotient map, there is a natural identification for any  $1 \leq$

<sup>13</sup>A metric space is said to be proper if any closed ball is compact.

<sup>14</sup>Necessarily,  $G$  must be unimodular if it has any lattice at all.



$q \leq \infty$

$$L^q(X) \cong L^q(Y)^K$$

where the right hand side is the space of  $K$  invariant functions.

To be consistent with the right action by inverses, for a subset  $F \subseteq G$  for any  $y \in Y$  we define  $yF := \{yf^{-1} \mid f \in F\}$ .

There is a residual action of the open subgroup  $N_G(K)$  on  $X$ . In particular, for any identity neighborhood  $B \subset N_G(K)$  and for all  $x \in X$  the set  $xB$  is an open neighborhood of  $x$ . Moreover, these sets form a base for the topology of  $X$  around  $x$ .

Fix  $a \in G$ . We are interested in the dynamics of the right action by  $a$  on  $Y$ . Obviously, if we want the induced averaging operator on  $X$  to have  $\|T\|_{L^2_0(X)} < 1$  the space  $X$  can not carry an  $a$  action, hence  $a$  can not belong to  $N_G(K)$ .

**4.1.2.  $S$ -Algebraic Groups.** All our applications will be in the following setting. Let  $\mathbf{G}$  be a linear reductive algebraic group defined over  $\mathbb{Q}$ . Fix  $S$  a finite set of places for  $\mathbb{Q}$  including at least one finite place. If  $\mathbf{G}(\mathbb{R})$  is not compact then  $S$  must include the infinite place.

Define  $G = G_S := \prod_{v \in S} \mathbf{G}(\mathbb{Q}_v)$  and let  $\Gamma < G_S$  be a congruence lattice. Let  $p \in S$  be a finite place and set  $K < \mathbf{G}(\mathbb{Q}_p)$  a compact open subgroup. We have  $N_{G_S}(K) \cong \prod_{v \in S \setminus \{p\}} \mathbf{G}(\mathbb{Q}_v) \times N_{\mathbf{G}(\mathbb{Q}_p)}(K)$  which is an open subgroup of  $G_S$ .

Not any subgroup  $K$  will define a state space with property (M). Nevertheless, we shall construct using the action of  $\mathbf{G}(\mathbb{Q}_p)$  on the associated affine Bruhat-Tits building natural examples of subgroups  $K$  which exhibit property (M) for the right action by  $a$ .

**4.2. Criteria for Property (M).** We now discuss what properties of  $K$  and  $a$  insures that the dynamical system given by the right action of  $a$  on  $Y$  has property (M) with respect to the state space  $X$ .

**Definition 4.1.** For each  $s, t \in \mathbb{Z}$ ,  $t \geq s$ , define

$$K^{(s,t)} = \bigcap_{s \leq i \leq t} a^{-i} K a^i$$

and

$$K^{(-\infty, t)} := \bigcap_{-\infty < i \leq t} a^{-i} K a^i$$

This is a compact subgroup of  $G$ .

Notice that if  $s' \leq s \leq t \leq t'$  then  $K^{(s', t')} \leq K^{(s, t)}$ . Moreover, the orthogonal projection  $L^2(X)^{K^{(s', t')}} \rightarrow L^2(X)^{K^{(s, t)}}$  is given by

$$f \mapsto f^{K^{(s, t)}}(x) := \int_{K^{(s, t)}} f(xk^{-1}) dk \text{ for } \mu\text{-almost every } x \in X$$

where we integrate with respect to the unique probability Haar measure on a compact subgroup.

Denote by  $U: L^2(Y) \rightarrow L^2(Y)$  the Koopman operator associated to the right action by  $a$  on  $Y$ . A computation shows that  $U \left( L^2(X)^{K^{(s,t)}} \right) \subseteq L^2(X)^{K^{(s+1,t+1)}} \subseteq L^2(X)^{K^{(s,t+1)}}$ .

**Lemma 4.2.** *Suppose that for all  $n \in \mathbb{N}$  the following diagram commutes*

$$\begin{array}{ccc} L^2(Y)^{K^{(0,n)}} & \xrightarrow{U} & L^2(Y)^{K^{(0,n+1)}} \\ \downarrow & & \downarrow \\ L^2(Y)^{K^{(0,n-1)}} & \xrightarrow{U} & L^2(Y)^{K^{(0,n)}} \end{array}$$

*Then the system  $Y$  with the right  $a$ -action has property (M) with respect to the state space  $X$ .*

*Proof.* First we show that  $P_X U P_X f = P_X U f$  for all  $f \in L^2(Y)^{K^{(0,n)}}$  and for all  $n \geq 0$ . This is proven by induction. The case  $n = 0$  is trivial. The induction step from  $n - 1$  to  $n$  follows from the complete commutativity of the following diagram.

$$\begin{array}{ccccc} & L^2(Y)^{K^{(0,n)}} & \xrightarrow{U} & L^2(Y)^{K^{(0,n+1)}} & \\ & \downarrow & & \downarrow & \\ P_X \left( \begin{array}{ccc} L^2(Y)^{K^{(0,n-1)}} & \xrightarrow{U} & L^2(Y)^{K^{(0,n)}} \\ \downarrow P_X & & \downarrow P_X \\ L^2(Y)^K & \xrightarrow{U} & L^2(Y)^{K^{(0,1)}} \end{array} \right) & \xrightarrow{P_X} & L^2(Y)^K & \leftarrow P_X \end{array}$$

The commutativity of the lower square is the induction assumption and the commutativity of the upper square is the hypothesis of the lemma.

The claim we have just proven implies that  $P_X U P_X f = P_X U f$  for all  $f \in \bigoplus_{n=0}^k L^2(Y)^{K^{(0,n)}}$  for all  $k \in \mathbb{N}$ . Using the continuity of  $P_X$  and  $U$  we see that  $P_X U P_X f = P_X U f$  for all  $f \in \mathcal{M}_a := \widehat{\bigoplus_{n=0}^{\infty} L^2(Y)^{K^{(0,n)}}}$ . The subspace  $\mathcal{M}_a$  is a closed  $U$ -stable  $L^\infty(X)$ -module that satisfies the requirements for property (M).  $\square$

**Corollary 4.3.** *Assume that there are  $\omega_1, \dots, \omega_k \in K^{(-\infty, 0)}$  such that*

$$K = \prod_{j=1}^k \omega_j K^{(0,1)}$$

*Then the system  $Y$  with the right  $a$ -action has property (M) with respect to the state space  $X$ .*

*Proof.* We claim that for any integer  $n \geq 0$

$$(9) \quad K^{(-n,0)} = \prod_{j=1}^k \omega_j K^{(-n,1)}$$

We prove this by induction. The case  $n = 0$  is just the hypothesis of the corollary. We assume the claim for  $n - 1$  and prove it for  $n$

$$\begin{aligned} K^{(-n,0)} &= a^n K a^{-n} \cap K^{(-(n-1),0)} = a^n K a^{-n} \cap \prod_{j=1}^k \omega_j K^{(-(n-1),1)} \\ &= \prod_{j=1}^k \left( a^n K a^{-n} \cap \omega_j K^{(-(n-1),1)} \right) = \prod_{j=1}^k \omega_j \left( a^n K a^{-n} \cap K^{(-(n-1),1)} \right) \\ &= \prod_{j=1}^k \omega_j K^{(-n,1)} \end{aligned}$$

In the second line we have used that  $\omega_j \in a^n K a^{-n}$  for all  $j$ .

We conjugate equation (9) by  $a^{-(n-1)}$  and have

$$K^{(0,n-1)} = \prod_{j=1}^k \left( a^{-(n-1)} \omega_j a^{n-1} \right) K^{(0,n)}$$

We can now check that the diagram of Lemma 4.2 does commute. The projection  $L^2(Y)^{K^{(0,n)}} \rightarrow L^2(Y)^{K^{(0,n-1)}}$  is given for any  $n$  by

$$f(x) \mapsto \frac{1}{k} \sum_{j=1}^k f \left( x a^{-(n-1)} \omega_j a^{n-1} \right)$$

Applying this formula for  $n$  and  $n + 1$  shows that the required diagram commute by a direct computation.  $\square$

*Remark 4.4* (Transferring property (M)). Notice that the hypothesis of Corollary 4.3 depends only on the subgroups  $K^{(s,t)}$  for  $s \leq t$  integers. If we prove that  $K < G$  and  $a \in G$  satisfy the conditions of Corollary 4.3 then for any other  $K' < G'$  and  $a' \in G'$  such that  $K'^{(s,t)} \cong K^{(s,t)}$  for all  $s \leq t$  we have property (M) for the spaces associated to  $G'$ ,  $K'$  with the right action by  $a'$ .

This allows us to verify property (M) on a complicated space by studying a simpler one.

**4.3. Walks on Buildings.** Let now  $G$  be a group acting by simplicial automorphisms on a locally finite thick affine building  $\Delta$ . Denote by  $G^\bullet$  the subgroup of type-preserving transformations. Assume that  $G^\bullet$  acts strongly transitively on the building  $\Delta$ , that is the action is transitive on pairs  $(\mathcal{C}, \mathcal{A})$  of a chamber  $\mathcal{C}$  and an apartment  $\mathcal{A}$  containing  $\mathcal{C}$ .

In this section we construct subgroups of  $G$  for which the hypothesis of Corollary 4.3 holds with respect to the action by a specific element

$a \in G$ . This generalizes non-backtracking walks on the Bruhat-Tits tree of  $\mathbf{PGL}_2(\mathbb{Q}_p)$ .

Fix an apartment  $\mathcal{A} \subset \Delta$ . Let  $N < G$  be the stabilizer of the apartment and denote by  $N_{\text{trans}}$  the elements acting on the apartment  $\mathcal{A}$  by translations. Fix  $e \neq a \in N_{\text{trans}}$ . We are interested in the action of this  $a$ .

Fix a special vertex  $v_0 \in \mathcal{A}$ . Let  $v_i = a^i.v_0$  for all  $i \in \mathbb{Z}$ . Assume that  $v_0$  and  $v_1$  *do not* share a common wall. This is a regularity condition on  $a$ .

Let  $\mathcal{S}^+ \subset \mathcal{A}$  be a sector in the apartment  $\mathcal{A}$  emanating from  $v_0$  and including  $v_1$ . This sector is unique. Let  $\mathcal{C}^+$  be the chamber in  $\mathcal{A}$  which includes  $v_0$  and lies at the base of  $\mathcal{S}^+$ .

Denote by  $\mathcal{S}^-$  the unique sector emanating from  $v_0$  and including  $\mathcal{C}^+$ . As  $v_0$  and  $v_1$  do not share a common wall the sectors  $\mathcal{S}^+$  and  $\mathcal{S}^-$  have opposite orientations. Let  $\mathcal{C}^-$  be the chamber lying at the base of  $\mathcal{S}^-$ .

Set  $I^+ < G$  and  $I^- < G$  to be the Iwahori subgroups fixing  $\mathcal{C}^+$  and  $\mathcal{C}^-$  respectively. The subgroup we study is the *arrow* subgroup

$$K_{\rightarrow} = I^+ \cap I^-$$

If the  $v_0$  and  $v_1$  belong to the same chamber then actually  $\mathcal{C}^+ = \mathcal{C}^-$  and  $K_{\rightarrow}$  is an Iwahori subgroup.

If  $\mathcal{C}^+ \neq \mathcal{C}^-$  then the arrow subgroup stabilizes more than just the two chambers  $\mathcal{C}^+, \mathcal{C}^-$ .

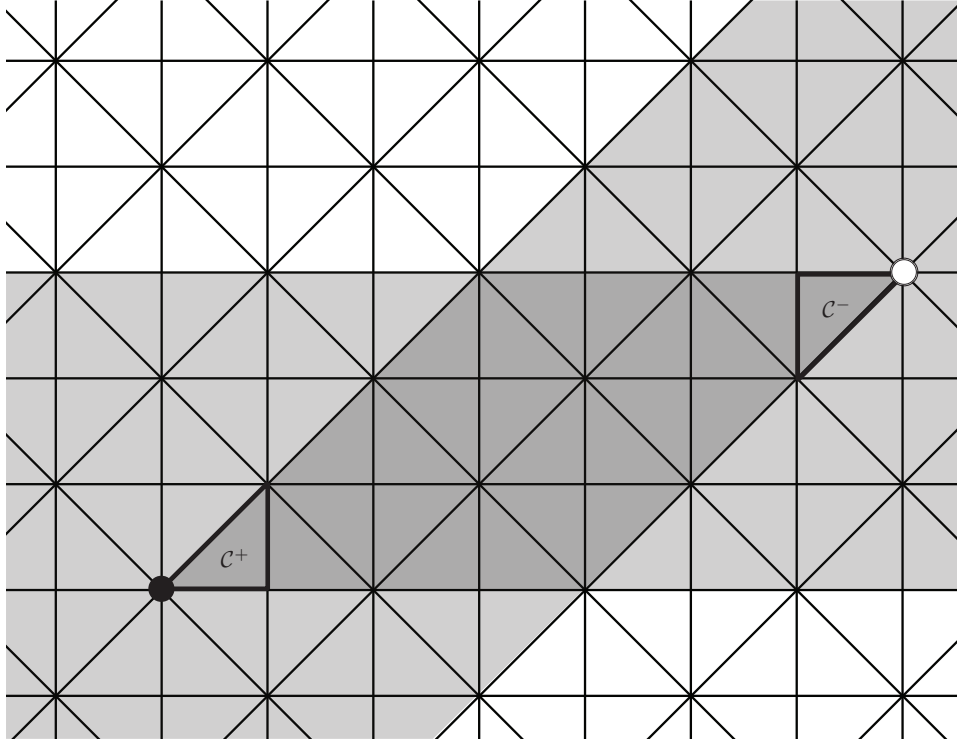
**Definition 4.5.** Let  $\mathcal{A}_0$  be an apartment in the building  $\Delta$  and let  $w_0, w_1 \in \mathcal{A}_0$  be two special vertices in the apartment which do not lie on a common wall. Let  $\mathcal{S}_0$  be the unique sector in  $\mathcal{A}_0$  emanating from  $w_0$  and including  $w_1$ . Let  $\mathcal{S}_1$  be the sector emanating from  $w_1$  with opposite orientation to  $\mathcal{S}_0$ . We call the intersection  $\mathcal{R} = \mathcal{S}_0 \cap \mathcal{S}_1$  a *rhomboid* with corners  $w_0$  and  $w_1$ . It is a bounded convex subset in the apartment  $\mathcal{A}_0$  containing  $w_0$  and  $w_1$ .

**Proposition 4.6.** *The arrow subgroup  $K_{\rightarrow}$  stabilizes pointwise the whole rhomboid  $\mathcal{R}$  with corners  $v_0$  and  $v_1$ .*

*Proof.* Every minimal gallery  $\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_l$  connecting  $\mathcal{C}^+$  and  $\mathcal{C}^-$  is contained in the apartment  $\mathcal{A}$ , see [Gar97, §4.5]. An element  $k \in K_{\rightarrow}$  stabilizes pointwise the chambers  $\mathcal{C}^+$  and  $\mathcal{C}^-$ , hence  $k.\mathcal{C}_0, k.\mathcal{C}_1, \dots, k.\mathcal{C}_l$  is also contained in  $\mathcal{A}$ . Because an apartment is a thin complex and  $k$  stabilizes the first and last chamber of the gallery it must stabilize pointwise the whole minimal gallery.

Every chamber  $\mathcal{C}_0 \subset \mathcal{R}$  belongs to some minimal gallery connecting  $\mathcal{C}^+$  and  $\mathcal{C}^-$ , see [Gar97, §16.1], hence it is stabilized by  $K_{\rightarrow}$ . □

**Proposition 4.7.** *The hypothesis of Corollary 4.3 holds for the arrow subgroup  $K_{\rightarrow}$  and the right action by  $a$ . Specifically, there exists  $\omega_1, \dots, \omega_k \in$*

FIGURE 1. An apartment in a  $\widetilde{C}_2$  building.

$K_{\rightarrow}^{(-\infty, 0)}$  such that

$$K_{\rightarrow} = \prod_{j=1}^k \omega_j K_{\rightarrow}^{(0, 1)}$$

*Proof.* Denote by  $\mathcal{P} := \mathcal{C}^+ \cup \mathcal{C}^-$  the pair of chambers defining  $K_{\rightarrow}$ . The group  $K_{\rightarrow}^{(0, 1)}$  stabilizes pointwise  $\mathcal{P}$  and  $a.\mathcal{P}$ .

Let  $\omega K_{\rightarrow}^{(0, 1)}$  be a right coset in  $K_{\rightarrow}$  then

$$\omega K_{\rightarrow}^{(0, 1)} = \{g \in G \mid g.\mathcal{P} = \mathcal{P} \text{ and } ga.\mathcal{P} = \omega a.\mathcal{P}\}$$

Recall that  $v_1 = a.v_0$ . The vertex  $v_1$  is stabilized by  $\omega K_{\rightarrow}^{(0, 1)}$ , hence the chamber  $\omega a.\mathcal{C}^+$  necessarily has  $v_1$  as a vertex. Moreover the metric distance between  $\omega a.\mathcal{C}^-$  and  $v_1$  is equal to the distance between  $a.\mathcal{C}^-$  and  $v_1$  because the vertex  $v_1$  is fixed by  $\omega$ . Hence there is a ball of fixed finite radius around  $v_1$  containing *all* the chamber pairs  $\omega a.\mathcal{P}$  coming from all the right cosets of  $K_{\rightarrow}^{(0, 1)}$  in  $K_{\rightarrow}$ . By the local finiteness of the building, there are only finitely many possibilities for such pairs of chambers and hence only finitely many right cosets of  $K_{\rightarrow}^{(0, 1)}$  in  $K_{\rightarrow}$ .

The chamber pair  $a.\mathcal{P}$  is contained in the sector  $a.\mathcal{S}^+$  and the chamber pair  $\mathcal{P}$  is contained in the sector  $\mathcal{S}^-$ . Thus for any  $\omega \in K_{\rightarrow}$  the pair of chambers  $\omega a.\mathcal{P}$  is contained in the sector  $\omega a.\mathcal{S}^+$ .

Our plan is to find  $\omega' \in G^\bullet$  which stabilizes pointwise  $\mathcal{S}^-$  and sends  $a.\mathcal{S}^+$  to  $\omega a.\mathcal{S}^+$  and than show that such  $\omega'$  must be contained in  $\omega K_{\rightarrow}^{(0,1)} \cap K_{\rightarrow}^{(-\infty,0)}$ . This  $\omega'$  will be the required representative of the coset and this will finish the proof.

Fix  $\omega \in K_{\rightarrow} \setminus K_{\rightarrow}^{(0,1)}$  for the rest of the proof <sup>15</sup>.

Let  $W_{\text{aff}}$  the affine Weyl group of the building  $\Delta$ . We shall make use of the  $W_{\text{aff}}$  valued distance function on chambers in  $\Delta$ , see [Gar97, §15.3]. For every two chamber  $\mathcal{C}, \mathcal{D} \in \Delta$  we have  $\delta(\mathcal{C}, \mathcal{D}) \in W_{\text{aff}}$ . If  $\mathcal{A}_0$  is any apartment containing  $\mathcal{C}$  and  $\mathcal{D}$  we have a canonical action of  $W_{\text{aff}}$  on  $\mathcal{A}_0$  and <sup>16</sup>  $\delta(\mathcal{C}, \mathcal{D}).\mathcal{D} = \mathcal{C}$ . Moreover, for any  $g \in G$  and any two chambers  $\mathcal{C}, \mathcal{D}$  we have  $\delta(g.\mathcal{C}, g.\mathcal{D}) = \delta(\mathcal{C}, \mathcal{D})$ .

Let  $\bar{w} \in W_{\text{sph}}$  be the longest element of the spherical Weyl group. We immediately see that

$$(10) \quad \delta(\mathcal{C}^-, \omega a.\mathcal{C}^+) = \delta(\omega.\mathcal{C}^-, \omega a.\mathcal{C}^+) = \delta(\mathcal{C}^-, a.\mathcal{C}^+) = \bar{w}$$

We wish to understand  $\omega a.\mathcal{S}^+ \cap \mathcal{S}^-$ . Obviously, the vertex  $v_1$  belongs to the intersection, we claim that this all the intersection. The set  $\omega a.\mathcal{S}^+ \cap \mathcal{S}^-$  lies in the intersection of the two apartments  $\mathcal{A} \cap \omega.\mathcal{A}$ . The intersection of the two apartments is a convex subset in  $\mathcal{A}$  including  $v_1$ . Hence if  $\omega a.\mathcal{S}^+ \cap \mathcal{S}^-$  is strictly larger then  $\{v_1\}$  then by convexity the chambers  $\omega a.\mathcal{C}^+$  and  $\mathcal{C}^-$  share a common face of dimension greater then 1, this contradicts the fact  $\delta(\mathcal{C}^-, \omega a.\mathcal{C}^+) = \bar{w}$ . We have thus proven  $\omega a.\mathcal{S}^+ \cap \mathcal{S}^- = \{v_1\}$ .

We now construct  $\omega' \in G$  such that  $\omega'.\mathcal{S}^- = \mathcal{S}^-$  and  $\omega'a.\mathcal{S}^+ = \omega a.\mathcal{S}^+$ . Look at the subset  $\mathcal{D}_\omega := \mathcal{S}^- \cup \omega a.\mathcal{S}^+$ . We claim that this subset is *strongly isometric* to  $\mathcal{D} := \mathcal{S}^- \cup a.\mathcal{S}^+ \subset \mathcal{A}$ . Recall that two subsets of chambers,  $\mathcal{E}$  and  $\mathcal{F}$ , are strongly isometric if there exists a bijection  $f: \mathcal{E} \rightarrow \mathcal{F}$  such that for any two chambers  $\mathcal{C}, \mathcal{D} \in \mathcal{E}$  we have  $\delta(f(\mathcal{C}), f(\mathcal{D})) = \delta(\mathcal{C}, \mathcal{D})$ .

We define the strong isometry  $f: \mathcal{D} \rightarrow \mathcal{D}_\omega$  by

$$f(\mathcal{C}) = \begin{cases} \mathcal{C} & \text{if } \mathcal{C} \in \mathcal{S}^- \\ \omega.\mathcal{C} & \text{if } \mathcal{C} \in a.\mathcal{S}^+ \end{cases}$$

This is well defined because the intersection of  $\mathcal{S}^-$  and  $\omega a.\mathcal{S}^+$  includes only the vertex  $v_1$ . It is obvious that  $f$  preserves  $\delta$  between two chamber which are both contained either in  $\mathcal{S}^-$  or in  $a.\mathcal{S}^+$ . Let now  $\mathcal{C} \in \mathcal{S}^-$  and  $\mathcal{C}' \in a.\mathcal{S}^+$ .

<sup>15</sup>For  $\omega \in K_{\rightarrow}^{(0,1)}$  there is nothing to prove as we can take the identity as a representative of  $\omega K_{\rightarrow}^{(0,1)}$ .

<sup>16</sup>To be consistent with the definitions of [Gar97, §15.3] we need to let  $\delta(\mathcal{C}, \mathcal{D})$  act on  $\mathcal{A}_0$  on the right after fixing an identification of  $\mathcal{A}_0$  with  $W_{\text{aff}}$ .

We use the multiplicative property of  $\delta$  and have

$$\begin{aligned}\delta(f(\mathcal{C}), f(\mathcal{C}')) &= \delta(\mathcal{C}, \omega.\mathcal{C}') = \delta(\mathcal{C}, \mathcal{C}^-)\delta(\mathcal{C}^-, \omega a.\mathcal{C}^+)\delta(\omega a.\mathcal{C}^+, \omega.\mathcal{C}') \\ &= \delta(\mathcal{C}, \mathcal{C}^-)\delta(\mathcal{C}^-, a.\mathcal{C}^+)\delta(a.\mathcal{C}^+, \mathcal{C}') \\ &= \delta(\mathcal{C}, \mathcal{C}')\end{aligned}$$

One can show using an analogous calculation that  $\delta(f(\mathcal{C}'), f(\mathcal{C})) = \delta(\mathcal{C}', \mathcal{C})$ .

We have shown that  $\mathcal{D}_\omega$  is strongly isometric to a subset of an apartment, hence there exists an apartment  $\mathcal{A}_\omega$  such that  $\mathcal{D}_\omega \subset \mathcal{A}_\omega$ , see [Gar97, §15.5]. Notice that  $\mathcal{C}^- \subset \mathcal{S}^- \subset \mathcal{A}_\omega$ . We now use the strong transitivity of  $G^\bullet$  acting on  $\Delta$  to deduce that there is  $\omega' \in G^\bullet$  sending the pair  $(\mathcal{C}^-, \mathcal{A})$  to  $(\mathcal{C}^-, \mathcal{A}_\omega)$ .

This  $\omega'$  must send  $\mathcal{S}^-$  to the sector emanating from  $\omega'(v_1) = v_1$  in the orientation defined by  $\omega'(\mathcal{C}^-) = \mathcal{C}^-$  which is exactly  $\mathcal{S}^-$ . In particular, as sectors are thin and  $\omega'$  preserves types we conclude that  $\omega'$  fixes  $\mathcal{S}^-$  pointwise.

Moreover, we have

$$\delta(\mathcal{C}^-, \omega' a.\mathcal{C}^+) = \delta(\omega' \mathcal{C}^-, \omega' a.\mathcal{C}^+) = \delta(\mathcal{C}^-, a.\mathcal{C}^+) = \bar{w}$$

Also  $\delta(\mathcal{C}^-, \omega a.\mathcal{C}^+) = \bar{w}$ , but both  $\omega a.\mathcal{C}^+$  and  $\omega' a.\mathcal{C}^+$  belong to the apartment  $\mathcal{A}_\omega$  and have the same  $\delta$  distance from  $\mathcal{C}^- \in \mathcal{A}_\omega$  hence  $\omega' a.\mathcal{C}^+ = \omega a.\mathcal{C}^+$ . We see that  $\omega' a.\mathcal{S}^+$  is a sector in the same apartment  $\mathcal{A}_\omega$  as  $\omega a.\mathcal{S}^+$  with the same base point  $v_1$  and the same chamber at base, hence  $\omega' a.\mathcal{S}^+ = \omega a.\mathcal{S}^+$  as required.

We are left only with proving that  $\omega' \in \omega K_{\rightarrow}^{(0,1)} \cap K_{\rightarrow}^{(-\infty,0)}$ . Because  $\omega'$  stabilizes pointwise  $\mathcal{S}^-$ , to prove that  $\omega' \in K_{\rightarrow}^{(-\infty,0)}$  it is enough to show that  $a^{-n}.\mathcal{P} \subset \mathcal{S}^-$  for all  $n \geq 0$ . We prove this claim by induction. For  $n = 0$  it follows from the definition of  $\mathcal{S}^-$ . Assume the claim for  $n$ :  $a^{-n}.\mathcal{P} \subset \mathcal{S}^-$ . By applying  $a^{-1}$  we have  $a^{-(n+1)}.\mathcal{P} \subset a^{-1}.\mathcal{S}^-$ . The sector  $a^{-1}.\mathcal{S}^-$  emanates from  $v_0 \in \mathcal{S}^-$  and because  $a^{-1}$  acts as a translation on the apartment  $\mathcal{A}$  we have that  $a^{-1}.\mathcal{S}^-$  has the same orientation in  $\mathcal{A}$  as  $\mathcal{S}^-$ . Hence  $a^{-1}.\mathcal{S}^- \subset \mathcal{S}^-$  and  $a^{-(n+1)}.\mathcal{P} \subset \mathcal{S}^-$  as required.

At last we show that  $\omega' \mathcal{P} = \omega \mathcal{P}$  which would imply  $\omega' K_{\rightarrow}^{(0,1)} = \omega K_{\rightarrow}^{(0,1)}$ . We already know that  $\omega' a.\mathcal{C}^+ = \omega a.\mathcal{C}^+$ , where the equality must be pointwise as both  $\omega'$  and  $\omega$  are type-preserving<sup>17</sup>. Now both  $\omega' a.\mathcal{C}^-$  and  $\omega a.\mathcal{C}^-$  belong to the apartment  $\mathcal{A}_\omega$  and  $\delta(\omega a.\mathcal{C}^+, \omega a.\mathcal{C}^-) = \delta(a.\mathcal{C}^+, a.\mathcal{C}^-) = \delta(\omega a.\mathcal{C}^+, \omega' a.\mathcal{C}^-)$ , so  $\omega' a.\mathcal{C}^- = \omega a.\mathcal{C}^-$  with pointwise equality.  $\square$

**4.4. Walks on  $S$ -Algebraic Quotients.** In this section we finally present the main class of systems with property (M) which we are interested in. Let  $\mathbf{G}$  be a reductive linear algebraic group defined over  $\mathbb{Q}$  and let  $S$  be a finite set of places of  $\mathbb{Q}$ . Set  $G_S := \prod_{v \in S} \mathbf{G}(\mathbb{Q}_v)$ .

We fix some finite place  $p \in S$  and let  $G_p := \mathbf{G}(\mathbb{Q}_p)$  act on the associated affine Bruhat-Tits building  $\Delta$ . The building  $\Delta$  has been constructed for  $\mathbf{SL}_n$

<sup>17</sup>The element  $\omega$  belongs to an Iwahori hence preserves types.

by Goldman and Iwahori [GI63], for split groups by Iwahori and Matsumoto [IM65] and generally by Bruhat and Tits [BT84].

**Proposition 4.8.** *Let  $\Gamma < G_S$  be some lattice and define  $Y := \Gamma \backslash G_S$ . Let  $a \in G_p$  be a semisimple element. Denote by  $A < G_p$  the maximal split subtorus<sup>18</sup> of a maximal torus containing  $a$ .*

*The element  $a^{-1}$  acts on the apartment in  $\Delta$  associated to  $A$ , assume that it sends a fixed special vertex  $v_0$  to a different special vertex  $v_1$  which does not have a shared wall with  $v_0$ . This assumption is fulfilled whenever  $a$  is  $\mathbb{Q}_p$ -regular and does not belong to a compact subgroup.*

*Let  $K_{\rightarrow} < G_p$  be the arrow subgroup corresponding to  $a^{-1}$  and the vertex  $v_0$  in the apartment associated to  $A$ . Define  $X := Y / K_{\rightarrow}$ . Then  $Y \rightarrow X$  has property (M) with respect to the right action by  $a$ .*

*Proof.* The group  $K_{\rightarrow}$ , being the intersection of two Iwahori subgroup, is a compact open subgroup of  $G_p$ . In particular, it is a compact subgroup of  $G_S$  and  $X$  is a Hausdorff quotient space.

The claim is a direct consequence of Corollary 4.3 and Proposition 4.7.  $\square$

## 5. APPLICATIONS TO EQUIDISTRIBUTION

We return to the general setting of §4.1.1. We study the dynamical system  $(Y, \mathcal{Y}, m, S)$ ,  $Y := \Gamma \backslash G$  where  $G$  is a second-countable locally compact topological group,  $\Gamma < G$  – a lattice,  $m$  – the probability Haar measure on  $Y$  and the transformation  $S$  is the right action by some fixed  $a \in G$ . The state space is  $X := Y / K$  where  $K < G$  is a compact subgroup such that  $N_G(K) < G$  is an open subgroup.

Moreover we denote the associated averaging operator by  $T$  and assume that  $\lambda := \|T\|_{L_0^2(X)} < 1$ .

For any  $a$ -invariant Borel probability measure  $\nu$  on  $Y$  we denote by  $h_\nu(a)$  the measure theoretic entropy of  $\nu$  with respect to the right  $a$ -action. In particular,  $h_m(a)$  is the entropy of the Haar measure.

### 5.1. Bowen Balls.

**Definition 5.1.** Fix  $a \in G$ . Let  $B \subseteq G$  be an identity neighborhood. For  $s, t \in \mathbb{Z}$  such that  $s \leq t$  we define the  $B^{(s,t)}$  Bowen ball to be

$$B^{(s,t)} := \bigcap_{s \leq i \leq t} a^{-i} B a^i$$

We say that  $B$  is  $(a, h)$ -homogeneous for some  $h \geq 0$  if there exists  $C_{B,h} > 0$  which depends only on  $h, B, G$  and  $a$  such that for any  $s \leq 0 \leq t$

$$m(B^{(s,t)}) \geq C_{B,h} \exp(-h(t - s + 1))$$

---

<sup>18</sup>The torus  $A$  is not necessarily a maximal torus in  $G_p$  as we have not assumed that  $G$  is split over  $\mathbb{Q}_p$ .



**Definition 5.2.** The Haar measure  $m$  is  $a$ -homogeneous if for all  $h > h_m(a)$  there is a base for the topology of  $G$  around the identity of  $(a, h)$ -homogeneous neighborhoods.

*Remark 5.3.* In important cases much stronger homogeneity properties for identity neighborhoods hold with respect to the Haar measure. Specifically, we might have a concrete value for  $C_{B,h}$  in terms of  $m(B)$  and perhaps even be able to set  $h = h_m(a)$ .

A stronger homogeneity assumption will produce a stronger effective equidistribution result in Theorem 5.10.

In Proposition B.1 we show that in the  $S$ -arithmetic setting the Haar measure is  $(a, h_m(a))$ -homogeneous.

**Definition 5.4.** Let  $\varphi: Y \rightarrow [0, 1]$  be  $\mathcal{X}$ -measurable. For any identity neighborhood  $B \subset N_G(K)$  we define

$$\begin{aligned}\varphi^B(y) &= \operatorname{ess\,sup}_{b \in B} \varphi(yb) = \operatorname{ess\,sup} \varphi(yB^{-1}) \\ \varphi_B(y) &= \operatorname{ess\,inf}_{b \in B} \varphi(yb) = \operatorname{ess\,sup} \varphi(yB^{-1})\end{aligned}$$

The assumption  $B \subset N_G(K)$  implies that the functions  $\varphi^B(y)$  and  $\varphi_B(y)$  are  $\mathcal{X}$ -measurable.

*Remark 5.5.* If  $G$  is an  $S$ -algebraic group with a compact part at infinity then one can look at neighborhoods  $B \subseteq K_0$  where  $K_0$  is a compact-open subgroup of  $G$ . In this case if  $\varphi$  is  $K_0$  invariant, i.e.  $K_0$ -smooth, then  $\varphi^B = \varphi_B = \varphi$ .

**Lemma 5.6.** Let  $\varphi: Y \rightarrow [0, 1]$  be  $\mathcal{X}$ -measurable and  $B \subset N_G(K)$  an identity neighborhood. Let  $s, t \in \mathbb{Z}$  such that  $s \leq t$ . For all  $y \in Y$  and  $z \in yB^{(s,t)}$

(a) if  $\varphi$  is lower-semicontinuous then

$$\sum_{i=s}^t \varphi^B(za^{-i}) \geq \sum_{i=s}^t \varphi(ya^{-i})$$

(b) if  $\varphi$  is upper-semicontinuous then

$$\sum_{i=s}^t \varphi_B(za^{-i}) \leq \sum_{i=s}^t \varphi(ya^{-i})$$

*Remark 5.7.* If  $\varphi$  is the characteristic function of an open set then it is lower-semicontinuous and if it is the characteristic function of a closed set it is upper-semicontinuous.

*Proof.* If  $z \in yB^{(s,t)}$  then for all  $s \leq i \leq t$  we have  $za^{-i} \in ya^{-i}B$ , i.e. there exists  $b \in B$  such that  $za^{-i} = ya^{-i}b^{-1} \Leftrightarrow za^{-i}b = ya^{-i}$ . This implies

$$\begin{aligned}\varphi^B(za^{-i}) &= \operatorname{ess\,sup}_{b \in B} \varphi(za^{-i}b) \geq \phi(ya^{-i}) \\ \varphi_B(za^{-i}) &= \operatorname{ess\,inf}_{b \in B} \varphi(za^{-i}b) \leq \phi(ya^{-i})\end{aligned}$$

In the inequalities on the right we have used lower-semicontinuity or upper-semicontinuity respectively.  $\square$

**Lemma 5.8.** *Let  $B \subset G$  be an identity neighborhood. If  $B_0 \subset G$  is an identity neighborhood such that  $B_0^{-1}B_0 \subseteq B$  then for all  $s, t \in \mathbb{Z}$  such that  $s \leq t$*

$$B_0^{(s,t)^{-1}} B_0^{(s,t)} \subseteq B^{(s,t)}$$

*Remark 5.9.* If  $B$  is a subgroup then  $B_0 = B$  satisfies the conditions of the lemma.

*Proof.* For any  $s \leq i \leq t$

$$(a^{-i}B_0a^i)^{-1}(a^{-i}B_0a^i) = a^{-i}B_0^{-1}B_0a^i \subseteq a^{-i}Ba^i$$

hence

$$\begin{aligned} B_0^{(s,t)^{-1}} B_0^{(s,t)} &= \bigcap_{s \leq i \leq t} (a^{-i}B_0a^i)^{-1} \cdot \bigcap_{s \leq j \leq t} (a^{-j}B_0a^j) \\ &\subseteq \bigcap_{s \leq i \leq t} (a^{-i}B_0a^i)^{-1} (a^{-i}B_0a^i) \\ &\subseteq \bigcap_{s \leq i \leq t} (a^{-i}Ba^i) = B^{(s,t)} \end{aligned}$$

$\square$

**5.2. Effective Equidistribution.** The idea to use large deviation estimates to prove equidistribution of measures with good separation properties goes back to Linnik [Lin68], who has already proved an effective version of his result. It has been developed further by Ellenberg, Michel and Venkatesh [EMV13]. The novelty of our result is that it applies to measures whose separation is sub-optimal.

The relevant separation property for the large deviations methods is similar to the separation property used in showing that the limit measure has positive entropy [ELMV09, Proposition 3.3]. We require that the average over a large compact set  $\Omega$  of the measure of Bowen balls on a *single* scale  $n_0$  is exponentially small.

$$(11) \quad -\frac{1}{2n_0} \log \int_{\Omega} \mu \left( yB^{(-n_0, n_0)} \right) d\mu(y) \geq h$$

This should be contrasted with entropy, which by the Brin-Katok theorem [BK83], implies that the average over a large set of the logarithm of the measure of Bowen balls is small for all large enough scales,  $\forall n \geq N$ .

$$-\frac{1}{2n} \int_{\Omega} \log \mu \left( yB^{(-n, n)} \right) d\mu(y) \geq h_m(a) - \varepsilon$$

Obviously, these two properties are related but are not interchangeable. Entropy does not provide a single scale on which the average size of the Bowen balls is small, it rather provides a scale on which the average size of

the logarithm of the Bowen ball is small. For a fixed Bowen Ball  $B^{(-n_0, n_0)}$  this is a weaker property than (11) because  $(-\log)$  is a strictly convex function.

On the other hand, entropy requires information regarding size of Bowen balls on all large enough scales, while the separation property we provide is on a single scale. Specifically, the measure  $\mu$  might have zero entropy and our theorem can still provide non-trivial information. This is extremely useful in applications like [Lin68] and [EMV13].

**Theorem 5.10.** *Fix  $a \in G$  and assume that  $Y \rightarrow X$  has property (M) with respect to the right action by  $a$ . Fix further  $B \subset G$  an identity neighborhood and a compact subset  $\Omega \subset Y$ . Let  $\mu$  be an  $a$ -invariant Borel probability measure on  $Y$ . Fix two distinct integers  $s \leq 0 \leq t$  and let  $n = t - s + 1$ . Define*

$$\alpha := -\frac{1}{n} \log \int_{\Omega} \mu \left( y B^{(s, t)} \right) d\mu(y)$$

*Fix  $h > 0$  and let  $B_0 \subset N_G(K)$  be an identity neighborhood such that*

- (1)  $B_0$  is  $(a, h)$ -homogeneous,
- (2)  $B_0^{-1} B_0 \subseteq B$ ,
- (3) *the projection  $\pi: G \rightarrow Y$  is injective when restricted to  $y B_0$  for any  $y \in \Omega$ .*<sup>19</sup>

*For a  $\mathcal{X}$ -measurable function  $\varphi: Y \rightarrow [0, 1]$*

- (a) *if  $\varphi$  is lower-semicontinuous and  $1 > \mu(\varphi) \geq m(\varphi^{B_0})$  then for any*

$$\mu(\Omega) > \kappa \geq \frac{1 - \mu(\varphi)}{1 - m(\varphi^{B_0})}; \quad \lambda^{\mathbb{N}} \ni \theta \leq \frac{[\mu(\varphi) - (1 - \kappa)] / \kappa - m(\varphi^{B_0})}{1 - m(\varphi^{B_0})}$$

*we have*

$$\begin{aligned} & \frac{D([\mu(\varphi) - (1 - \kappa)] / \kappa \parallel \theta + (1 - \theta)m(\varphi^{B_0}))}{-\log \theta} \leq \frac{h - \alpha}{-\log \lambda} \\ & + \frac{1}{n} \left[ \log \left( \frac{\log \theta}{\log \lambda} \right) + D(\mu(\varphi) \parallel m(\varphi^{B_0})) \right. \\ & \left. - 2 \log \frac{\mu(\Omega) - \kappa}{2} - \log C_{B_0, h} \right] \end{aligned}$$

- (b) *if  $\varphi$  is upper-semicontinuous and  $0 < \mu(\varphi) \leq m(\varphi_{B_0})$  then for any*

$$\mu(\Omega) > \kappa \geq \frac{\mu(\varphi)}{m(\varphi_{B_0})}; \quad \lambda^{\mathbb{N}} \ni \theta \leq \frac{m(\varphi_{B_0}) - \mu(\varphi) / \kappa}{m(\varphi_{B_0})}$$

---

<sup>19</sup>Such  $B_0$  necessarily exists by the continuity and non vanishing of the injectivity radius.

we have

$$\begin{aligned} \frac{D(\mu(\varphi)/\kappa \parallel (1-\theta)m(\varphi^{B_0}))}{-\log \theta} &\leq \frac{h-\alpha}{-\log \lambda} \\ &+ \frac{1}{n} \left[ \log \left( \frac{\log \theta}{\log \lambda} \right) + D(\mu(\varphi) \parallel m(\varphi^{B_0})) \right. \\ &\left. - 2 \log \frac{\mu(\Omega) - \kappa}{2} - \log C_{B_0, h} \right] \end{aligned}$$

*Remark 5.11.* In the favorable case that  $Y$  is compact one can take  $\Omega = Y$  and deduce a simpler statement. Nevertheless, the statement for a general  $\Omega$  is useful for us even in compact spaces. This is because often it is easy to find a big compact set all whose points have some good typical behavior with respect to  $\mu$  but that is not the case for the whole space  $Y$ .

*Proof.* The idea of the proof is to use the bound on the average size of  $B$ -Bowen balls to generate many mutually disjoint  $B_0$ -Bowen balls around points which have typical statistics with respect to  $\mu$ . Then use the fact that  $B_0$  is an  $a$ -homogeneous neighborhood to show a lower bound on the mass of the union of these disjoint  $B_0$ -Bowen balls with respect to the *Haar* measure  $m$ . On the other hand, all the points in these  $B_0$ -Bowen balls behave typically with respect to  $\mu$ , at least on the scale of  $B_0$ . This allows us to derive an upper bound on the Haar measure of the union using our large deviation results in terms of  $m(\varphi^{B_0})$  or  $m(\varphi_{B_0})$ .

Let  $\kappa_B > 0$  be a constant to be optimized later and let  $\kappa > 0$  be a free variable whose range will be set later as well. For a function  $f: Y \rightarrow \mathbb{R}$  denote for any  $y \in Y$

$$A^{(s,t)} f(y) := \frac{1}{t-s+1} \sum_{i=s}^t f(ya^{-i})$$

Because  $m$  is  $a$  invariant we have  $\int A^{(s,t)} f d\mu = \int f d\mu$  for any integrable  $f$ .

*Points with good statistics and small Bowen balls.* We want first to find  $\mu$ -many points whose  $(s, t)$  orbit has statistics similar to  $\mu$ . Define

$$\begin{aligned} Y^\varphi &:= \left\{ y \in Y \mid A^{(s,t)} \varphi(y) > \frac{\mu(\varphi) - (1-\kappa)}{\kappa} \right\} \\ Y_\varphi &:= \left\{ y \in Y \mid A^{(s,t)} \varphi(y) < \frac{\mu(\varphi)}{\kappa} \right\} \end{aligned}$$

Then by Markov's inequality

$$\mu(Y^\varphi), \mu(Y_\varphi) \geq 1 - \kappa$$

Next we define  $\Omega_\varphi := Y_\varphi \cap \Omega$  and  $\Omega^\varphi := Y^\varphi \cap \Omega$ . Obviously,

$$\mu(\Omega^\varphi), \mu(\Omega_\varphi) \geq 1 - \kappa - \mu(\Omega^C) = \mu(\Omega) - \kappa$$

Necessarily, we need  $\kappa < \mu(\Omega)$  for this expression to be useful.

Next we look for  $\mu$ -many points for which the measure of the  $B^{(s,t)}$  Bowen ball around them is close to  $\exp(-\alpha n)$ . Define

$$\Omega_B := \left\{ y \in Y \mid \mu\left(yB^{(s,t)}\right) \leq \exp(-\alpha n)/\kappa_B \right\}$$

Again by Markov's inequality

$$\mu(\Omega_B) \geq 1 - \kappa_B$$

Finally, we have a bound on the  $\mu$ -measure of the sets with good statistics and small Bowen balls.

$$\mu(\Omega^\varphi \cap \Omega_B), \mu(\Omega_\varphi \cap \Omega_B) \geq \mu(\Omega) - \kappa - \kappa_B$$

*Disjoint Bowen balls with good statistics.* We present an inductive procedure that constructs finite sets of points  $\Sigma^\varphi, \Sigma_\varphi$  such that

- (a)  $\Sigma^\varphi \subseteq \Omega^\varphi$  and  $\Sigma_\varphi \subseteq \Omega_\varphi$ ,
- (b) for any two points  $y_1, y_2 \in \Sigma^\varphi$  or  $y_1, y_2 \in \Sigma_\varphi$  we have  $y_1 B_0^{(s,t)} \cap y_2 B_0^{(s,t)} = \emptyset$ ,
- (c)  $|\Sigma^\varphi|, |\Sigma_\varphi| \geq \kappa_B \frac{\mu(\Omega) - \kappa - \kappa_B}{\exp(-\alpha n)}$ .

Let  $\Xi_0 = \Omega^\varphi \cap \Omega_B$  or  $\Xi_0 = \Omega_\varphi \cap \Omega_B$ . The inductive procedure goes as follows. Notice that  $\mu(\Xi_0) \geq \mu(\Omega) - \kappa - \kappa_B$ .

- (1) *Inductive assumption:* we have chosen  $y_1, \dots, y_k$  in  $\Omega^\varphi \cap \Omega_B$  or  $\Omega_\varphi \cap \Omega_B$  respectively such that  $y_i \notin \bigcup_{j=1}^{i-1} y_j B^{(s,t)}$  for any  $1 \leq i \leq k$ . Moreover, there is a set  $\Xi_k \subseteq \Omega^\varphi \cap \Omega_B$  or  $\Xi_k \subseteq \Omega_\varphi \cap \Omega_B$  respectively such that  $\mu(\Xi_k) \geq \mu(\Omega) - \kappa - \kappa_B - k \exp(-\alpha n)/\kappa_B$  and for each  $1 \leq i \leq k$  we have  $y_i B^{(s,t)} \cap \Xi_k = \emptyset$ .
- (2) If  $\mu(\Xi_k) = 0$  finish the procedure.
- (3) If  $\mu(\Xi_k) > 0$  choose  $y_{k+1} \in \Xi_k$  and set  $\Xi_{k+1} := \Xi_k \setminus y_{k+1} B^{(s,t)}$ .
- (4) Obviously,  $y_{k+1} B^{(s,t)} \cap \Xi_{k+1} = \emptyset$ . We have  $\Xi_{k+1} \subseteq \Xi_k$  hence for all  $1 \leq i \leq k$  also  $y_i B^{(s,t)} \cap \Xi_{k+1} = \emptyset$ .
- (5) Notice that  $y_{k+1} \in \Xi_k$  and  $\Xi_k \cap \bigcup_{i=1}^k y_i B^{(s,t)} = \emptyset$ , thus  $y_{k+1} \notin \bigcup_{i=1}^k y_i B^{(s,t)}$ .
- (6)  $\mu(\Xi_{k+1}) \geq \mu(\Xi_k) - \mu(y_{k+1} B^{(s,t)}) \geq \mu(\Xi_k) - \exp(-\alpha n)/\kappa_B \geq \mu(\Omega) - \kappa - \kappa_B - (k+1) \exp(-\alpha n)/\kappa_B$ , where we have used the fact that  $y_{k+1} \in \Omega_B$ .
- (7) Items (4), (5) and (6) show that our choice of  $\Xi_{k+1}, y_{k+1}$  fulfill the inductive assumption.

The procedure terminates when  $0 = \mu(\Xi_k) \geq \mu(\Omega) - \kappa - \kappa_B - k \exp(-\alpha n)/\kappa_B$ , meaning when  $k \geq \kappa_B \frac{\mu(\Omega) - \kappa - \kappa_B}{\exp(-\alpha n)}$ . Set  $\Sigma^\varphi = \{y_1, \dots, y_k\}$  and  $\Sigma_\varphi = \{y_1, \dots, y_k\}$  for  $\Xi_0 = \Omega^\varphi \cap \Omega_B$  and  $\Xi_0 = \Omega_\varphi \cap \Omega_B$  appropriately.

For each  $1 \leq j < i \leq k$  we have  $y_j \notin y_i B^{(s,t)}$ . Recall Lemma 5.8

$$B_0^{(s,t)-1} B_0^{(s,t)} \subseteq B^{(s,t)}$$

This implies for  $y_i, y_j$  as above that  $y_i B_0^{(s,t)} \cap y_j B_0^{(s,t)} = \emptyset$ . We conclude that the sets  $\Sigma^\varphi$  and  $\Sigma_\varphi$  have all the properties (a), (b) and (c) from above.

*The  $m$ -measure of the union of Bowen balls.* Set

$$\mathcal{D}^\varphi = \coprod_{y \in \Sigma^\varphi} y B_0^{(s,t)}$$

$$\mathcal{D}_\varphi = \coprod_{y \in \Sigma_\varphi} y B_0^{(s,t)}$$

Both unions are disjoint ones because of property (b) above. Because  $B_0$  is an  $(a, h)$ -homogeneous neighborhood we have a lower bound on the  $m$ -measure

$$m(\mathcal{D}^\varphi), m(\mathcal{D}_\varphi) \geq \kappa_B \frac{\mu(\Omega) - \kappa - \kappa_B}{\exp(-\alpha n)} C_{B_0, h} \exp(-nh)$$

At this point we can already optimize  $\kappa_B$  to make this measure as large as possible by setting  $\kappa_B = \frac{\mu(\Omega) - \kappa}{2}$ . We then have

$$(12) \quad m(\mathcal{D}^\varphi), m(\mathcal{D}_\varphi) \geq \frac{(\mu(\Omega) - \kappa)^2}{4} C_{B_0, h} \exp(-n(h - \alpha))$$

Because  $\mathcal{D}^\varphi$  and  $\mathcal{D}_\varphi$  are union of  $B_0$ -Bowen balls of points in  $Y^\varphi$  and  $Y_\varphi$  respectively, we have by Lemma 5.6 that if  $\varphi$  is semicontinuous in the appropriate way then

$$\forall y \in \mathcal{D}^\varphi: A^{(s,t)} \varphi^{B_0}(y) \geq \frac{\mu(\varphi) - (1 - \kappa)}{\kappa}$$

$$\forall y \in \mathcal{D}_\varphi: A^{(s,t)} \varphi_{B_0}(y) \leq \frac{\mu(\varphi)}{\kappa}$$

We now apply Theorem 3.2 to  $\varphi^{B_0}$  and  $\varphi_{B_0}$  respectively.

(i) If  $\varphi$  is lower-semicontinuous and

$$\kappa \geq \frac{1 - \mu(\varphi)}{1 - m(\varphi^{B_0})}; \quad \theta \leq \frac{[\mu(\varphi) - (1 - \kappa)] / \kappa - m(\varphi^{B_0})}{1 - m(\varphi^{B_0})}$$

then

$$m(\mathcal{D}^\varphi) \leq \frac{\log \theta}{\log \lambda} \exp D([\mu(\varphi) - (1 - \kappa)] / \kappa \| m(\varphi^{B_0}))$$

$$\times \exp \left[ -n(-\log \lambda) \frac{D([\mu(\varphi) - (1 - \kappa)] / \kappa \| \theta + (1 - \theta)m(\varphi^{B_0}))}{-\log \theta} \right]$$

(ii) If  $\varphi$  is upper-semicontinuous and

$$\kappa \geq \frac{\mu(\varphi)}{m(\varphi^{B_0})}; \quad \theta \leq \frac{m(\varphi^{B_0}) - \mu(\varphi) / \kappa}{1 - m(\varphi^{B_0})}$$

then

$$\begin{aligned}
m(\mathcal{D}_\varphi) &\leq \frac{\log \theta}{\log \lambda} \exp D(\mu(\varphi)/\kappa \| m(\varphi_{B_0})) \\
&\quad \times \exp \left[ -n(-\log \lambda) \frac{D(\mu(\varphi)/\kappa \| (1-\theta)m(\varphi_{B_0}))}{-\log \theta} \right] \\
&\leq \frac{\log \theta}{\log \lambda} \exp D(\mu(\varphi) \| m(\varphi_{B_0})) \\
&\quad \times \exp \left[ -n(-\log \lambda) \frac{D(\mu(\varphi)/\kappa \| (1-\theta)m(\varphi_{B_0}))}{-\log \theta} \right]
\end{aligned}$$

Inequalities (i) and (ii) combined with (12) imply the theorem.  $\square$

## 6. EFFECTIVE RIGIDITY OF THE MEASURE OF MAXIMAL ENTROPY

In this section we discuss how an  $L_0^2$ -norm bound for the averaging operator implies that the push forward of a measure with very large entropy to a state space is close in weak-\* sense to the Haar measure.

We deduce this result from a proposition about limits of sequences of measures with small Bowen balls on finer and finer scales.

Note that our results *do not* imply directly the uniqueness of the measure of maximal entropy. They just imply that all the measures of maximal entropy have a unique pushforward to the state space.

Although our methods are strong enough to show non-escape of mass for sequences of measure in some favorable situations we do not discuss this in the current section. See §7 for results about non-escape of mass.

*Standing assumptions.* We recall our standing assumptions and expand them. Suppose  $G$  is a second-countable locally compact topological group with a lattice  $\Gamma < G$ . Set  $Y := \Gamma/G$ . Let  $m$  be the unique probability Haar measure on  $Y$ , fix  $a \in G$  and consider the dynamical system defined by the right  $a$  action on  $Y$ .

Let  $K < G$  be a compact subgroup such that  $N_G(K) < G$  is open and define the state space  $X := Y/K$ . Assume a norm gap for the averaging operator  $\lambda := \|T\|_{L_0^2(X)} < 1$ .

We assume *in addition* that  $Y \rightarrow X$  has property (M) and that the Haar measure is  $a$ -homogeneous, see Proposition B.1.

**Proposition 6.1** (Limits of measures invariant under a fixed element). *Fix  $a \in G$  and let  $\{\mu_i\}_{i=1}^\infty$  be a sequence of  $a$ -invariant probability measures on  $Y$  such that  $\mu_i$  converges to a probability measure  $\mu$  in the weak-\* topology. Let  $h > 0$  be a fixed real number.*

*Suppose that we have a family  $\mathcal{F}$  of compact subset of  $Y$  such that for any  $\varepsilon > 0$  there exists  $\Omega \in \mathcal{F}$  such that  $\limsup_{i \rightarrow \infty} \mu_i(\Omega) \geq 1 - \varepsilon$ .*

*Assume further that for any compact subset  $\Omega \in \mathcal{F}$  there is an identity neighborhood  $B \subset G$  and a sequence of positive integers  $t_i \rightarrow_{i \rightarrow \infty} \infty$  such*

that for all  $i \in \mathbb{N}$

$$(13) \quad \int_{\Omega} \mu_i \left( y B^{(-t_i, t_i)} \right) d\mu_i(y) \leq C_{\Omega} e^{-2ht_i}$$

for some  $C_{\Omega} > 0$ .

Let  $\varphi: Y \rightarrow [0, 1]$  be a  $\mathcal{X}$ -measurable function, if  $\varphi$  is continuous or it is a characteristic function of either a closed or an open subset of  $X$  then the following holds.

(a) If  $1 > \mu(\varphi) > m(\varphi)$  then for any  $\lambda^{\mathbb{N}} \ni \theta < \frac{\mu(\varphi) - m(\varphi)}{1 - m(\varphi)}$  we have

$$\frac{D(\mu(\varphi) \parallel \theta + (1 - \theta)m(\varphi))}{-\log \theta} \leq \frac{h_m(a) - h}{-\log \lambda}$$

(b) If  $0 < \mu(\varphi) < m(\varphi)$  then for any  $\lambda^{\mathbb{N}} \ni \theta < \frac{m(\varphi) - \mu(\varphi)}{m(\varphi)}$  we have

$$\frac{D(\mu(\varphi) \parallel (1 - \theta)m(\varphi))}{-\log \theta} \leq \frac{h_m(a) - h}{-\log \lambda}$$

*Remark 6.2.* This result generalizes Linnik's approach to equidistribution using large deviation [Lin68]. It should be compared with the method of [EMV13] which served as a base for our work. The results of [EMV13] essentially use a variation of the proposition above for finite state spaces,  $h = h_m(a)$  and  $\theta \rightarrow 0$ .

A similar result for  $h = h_m(a)$  and  $Y := \mathbf{PGL}_2(\mathbb{R}) \backslash \mathbf{PGL}_2(\mathbb{R})$  is proven in [ELMV12, Theorem 4.2]. This last result is different from ours in a few significant ways. Instead of using a norm gap it uses an effective description of the uniqueness of measure of maximal entropy; it is stated for an action by an element  $a$  in the real place, which our results do not cover; and it applies for the whole space  $Y$  and not just to a state space. Moreover, [ELMV12, Theorem 4.2] allows to control escape of mass, which our methods can also accomplish.

The downside of [ELMV12, Theorem 4.2] is that it applies only to  $h = h_m(a)$ . Our result is effective, it provides useful information in a limited range of  $h < h_m(a)$ . Moreover, one can give a rate of convergence in our result – that is essentially Theorem 5.10.

*Remark 6.3.* We can take  $\mathcal{F}$  to be the collection of *all* compact subsets of  $Y$ , this would require a bound on the average measure of Bowen balls in any compact subset.

We show that for any  $\varepsilon > 0$  there exists a compact subset  $\Omega \subset Y$  such that  $\limsup_{i \rightarrow \infty} \mu_i(\Omega) \geq 1 - \varepsilon$ . This follows by a standard argument. Specifically, take a sequence of closed balls,  $\{C_{r_k}\}_{k=1}^{\infty}$ , around a fixed point in  $Y$  with a strongly monotonic sequence of radii  $\{r_k\}_{k=1}^{\infty}$  converging to infinity. We then have that all the set  $\partial C_{r_k}$  are mutually disjoint, hence  $\sum_{k=1}^{\infty} \mu(\partial C_{r_k}) = \mu(\bigcup_{k=1}^{\infty} \partial C_{r_k}) < \infty$ . This implies that  $\mu(\partial C_{r_k}) \rightarrow_{k \rightarrow \infty} 0$ , on the other hand  $\mu(C_{r_k}) \rightarrow_{k \rightarrow \infty} 1$ , hence  $\mu(C_{r_k}^{\circ}) = \mu(C_{r_k}) - \mu(\partial C_{r_k}) \rightarrow_{k \rightarrow \infty} 1$ .



Taking  $k$  large enough we get the required compact set  $\Omega = C_{r_k}$ . We actually have a stronger inequality

$$\liminf_{i \rightarrow \infty} \mu_i(\Omega) \geq \liminf_{i \rightarrow \infty} \mu_i(\Omega^\circ) \geq \mu(\Omega^\circ) \geq 1 - \varepsilon$$

If we take  $\mathcal{F}$  to be the collection of all compact sets then our result bears similarities to [ELMV09, Proposition 3.3] which under mildly stronger assumption shows an entropy lower bound  $h_\mu(a) \geq h$ . In spite of the similarities, it seems one cannot simply deduce the proposition above and [ELMV09, Proposition 3.3] from each other. In particular, in the  $S$ -algebraic setting our proposition gives strong information for  $h$  close to  $h_m(a)$  and no information at all for small values of  $h$ . On the other hand, [ELMV09, Proposition 3.3] gives non-trivial information in the whole range  $0 < h \leq h_m(a)$ , yet this information is weaker than ours. If the measure  $\mu$  is invariant under additional elements of  $G$  commuting with  $a$ , then having positive entropy alone can imply very strong results regarding the measure using the measure rigidity results of [Lin06, EKL06, EL15].

*Proof.* We prove the theorem for the case  $\mu(\varphi) > m(\varphi)$ , the proof of the second case is analogous.

Assume first that  $\varphi$  is continuous and compactly supported.

For any  $\varepsilon > 0$  let  $\Omega \in \mathcal{F}$  be a compact set such that  $\liminf_{i \rightarrow \infty} \mu(\Omega) > 1 - \varepsilon$ . Let  $B$  and  $\{t_i\}_{i=1}^\infty$  the identity neighborhood and the integer sequence corresponding to  $\Omega$  in the hypothesis of the proposition. Set in accordance with Theorem 5.10

$$(14) \quad \alpha := -\frac{1}{2t_i} \log \int_{\Omega} \mu_i \left( y B^{(-t_i, t_i)} \right) d\mu_i(y) \geq h - \frac{1}{2t_i} \log(C_\Omega)$$

Fix  $h_0 > h_m(a)$ . We use the standing assumption that the Haar measure is  $a$ -homogeneous to generate a descending sequence of  $(a, h_0)$ -homogeneous identity neighborhood  $B_0^j \subset N_G(K)$  satisfying the conditions of Theorem 5.10 and such that  $\bigcap_{j=1}^\infty B_0^j = \{e\}$ .

As  $\varphi$  is continuous and compactly supported it is actually uniformly continuous, hence  $\varphi^{B_0^j} \rightarrow_{j \rightarrow \infty} \varphi$  pointwise. By the monotone convergence theorem  $\lim_{j \rightarrow \infty} m(\varphi^{B_0^j}) = m(\varphi)$ .

Recall that  $\mu_i(\varphi) \rightarrow_{i \rightarrow \infty} \mu(\varphi) \in [0, 1]$ . We are dealing with the case  $\mu(\varphi) > m(\varphi)$ , hence for all  $i$  and  $j$  large enough one has  $\mu_i(\varphi) > m(\varphi^{B_0^j})$ .

By Theorem 5.10 for any  $\kappa$  and  $\theta \in \lambda^{\mathbb{N}}$  satisfying the following

$$\mu_i(\Omega) > \kappa \geq \frac{1 - \mu_i(\varphi)}{1 - m(\varphi^{B_0^j})}; \quad \theta \leq \frac{[\mu_i(\varphi) - (1 - \kappa)] / \kappa - m(\varphi^{B_0^j})}{1 - m(\varphi^{B_0^j})}$$

and for all  $i$  and  $j$  large enough we have

$$\begin{aligned} \frac{D\left([\mu_i(\varphi) - (1 - \kappa)] / \kappa \parallel \theta + (1 - \theta)m(\varphi^{B_0^j})\right)}{-\log \theta} &\leq \frac{h_0 - h}{-\log \lambda} + \frac{1}{2t_i} \log(C_\Omega) \\ &+ \frac{1}{2t_i} \left[ \log\left(\frac{\log \theta}{\log \lambda}\right) D\left(\mu_i(\varphi) \parallel m(\varphi^{B_0^j})\right) \right. \\ &\left. - 2 \log \frac{\mu_i(\Omega) - \kappa}{2} - \log C_{B_0^j, h_0} \right] \end{aligned}$$

where we have also applied inequality (14).

Recall that  $\limsup_{i \rightarrow \infty} \mu_i(\Omega) \geq 1 - \varepsilon$ . Now let  $i \rightarrow \infty$  along a subsequence attaining the limsup in the inequality for  $\Omega$ , then for any  $j$  large enough and any  $\kappa$  and  $\theta \in \lambda^{\mathbb{N}}$  satisfying

$$1 - \varepsilon > \kappa > \frac{1 - \mu(\varphi)}{1 - m(\varphi^{B_0^j})}; \quad \theta < \frac{[\mu(\varphi) - (1 - \kappa)] / \kappa - m(\varphi^{B_0^j})}{1 - m(\varphi^{B_0^j})}$$

we have

$$\frac{D\left([\mu(\varphi) - (1 - \kappa)] / \kappa \parallel \theta + (1 - \theta)m(\varphi^{B_0^j})\right)}{-\log \theta} \leq \frac{h_0 - h}{-\log \lambda}$$

The claim for  $\varphi$  continuous and compactly supported follows by first taking the limits  $j \rightarrow \infty$ ,  $\varepsilon \rightarrow 0$ ,  $\kappa \rightarrow 1$ , and at last  $h_0 \rightarrow h_m(a)$ .

For  $\varphi$  continuous not necessarily compactly supported one can take a monotonic sequence of compactly supported functions  $0 \leq \varphi_k \leq \varphi$  which converge pointwise to  $\varphi$ . Then the claim for  $\varphi$  follows from the claim for the functions  $\varphi_k$  by taking the limit  $k \rightarrow \infty$  and using the monotone convergence theorem.

If  $\varphi$  is the characteristic of an open or a closed subset of  $X$  then similarly it can be approximated by a sequence of continuous functions  $\varphi_k: X \rightarrow [0, 1]$  converging pointwise to  $\varphi$ . By Lebesgue's dominated convergence theorem for every measure  $\nu$  on  $Y$  one has  $\lim_{k \rightarrow \infty} \int_Y \varphi_k d\nu \rightarrow \int_Y \varphi d\nu$ . The result now follows from the claim for continuous functions by taking the limit  $k \rightarrow \infty$ .  $\square$

To pass from average mass of Bowen balls on a prescribed scale to a result regarding the classical notion of entropy we require the following standard lemma.

**Lemma 6.4.** *Let  $\mu$  be an  $a$ -invariant ergodic measure on  $Y$ . Then for every  $\varepsilon > 0$  there exists a compact subset  $\Omega \subseteq Y$ , with  $\mu(\Omega) > 1 - \varepsilon$ , and an identity neighborhood  $B \subseteq G$  such that for all  $t > N_\Omega \in \mathbb{N}$  and all  $y \in \Omega$*

$$\mu(yB^{(-t, t)}) \leq e^{-2t(h_\mu(a) - \varepsilon)}$$

*Proof.* This follows from an extension of the Brin-Katok theorem [BK83] to a second countable metric space, Theorem A.1.

Fix a left invariant metric  $d$  on  $G$  and let  $\tilde{d}$  be the induced metric on  $Y$ . Denote by  $\tilde{B}_r(y)$  be the open  $\tilde{d}$ -ball of radius  $r > 0$  centered at  $y \in Y$ . Let  $B_r \subset G$  be the ball of radius  $r > 0$  around the identity. We claim that for all  $y = \Gamma g \in Y$  and  $r > 0$  we have  $\tilde{B}_r(y) = yB_r$ . To see that we unwind the definition of  $\tilde{d}$

$$\begin{aligned} \tilde{d}(\Gamma g', \Gamma g) < r &\Leftrightarrow (\exists \gamma \in \Gamma) d(\gamma g', g) < r \Leftrightarrow (\exists \gamma \in \Gamma) d(g^{-1}\gamma g', e) < r \\ &\Leftrightarrow (\exists \gamma \in \Gamma) g^{-1}\gamma g \in B_r \Leftrightarrow (\exists \gamma \in \Gamma) \gamma g' \in gB_r \\ &\Leftrightarrow \Gamma g' \in \Gamma g B_r \end{aligned}$$

For any  $s \leq 0 \leq t$  define

$$\tilde{B}_r^{(s,t)}(y) := \left\{ y' \in Y \mid \forall s \leq n \leq t: \tilde{d}(y'a^{-n}, ya^{-n}) < r \right\}$$

By Theorem A.1 for almost all  $y \in Y$

$$\lim_{r \rightarrow 0} \liminf_{t \rightarrow \infty} \frac{-\log \mu(\tilde{B}_r^{(-t,t)}(y))}{2t} \geq h_\mu(a)$$

This implies by a standard argument that for every  $\varepsilon > 0$  there exists a compact subset  $\Omega \subseteq Y$  with  $\mu(\Omega) > 1 - \varepsilon$ , a radius  $r_\varepsilon > 0$  and an integer  $N_{r_\varepsilon} \in \mathbb{N}$  such that for all  $t \geq N_{r_\varepsilon}$ , for all  $r < r_\varepsilon$  and all  $y \in \Omega$

$$\frac{-\log \mu(\tilde{B}_r^{(-t,t)}(y))}{2t} \geq h_\mu(a) - \varepsilon$$

Fix  $B = B_{r_\varepsilon/2}$ . For any  $y \in Y$  we have  $yB^{(-t,t)} \subseteq \tilde{B}_{r_\varepsilon/2}^{(-t,t)}(y)$ . Set  $N_\Omega := N_{r_\varepsilon/2}$ .  $\square$

**Corollary 6.5.** *Let  $\mu$  be an  $a$ -invariant ergodic probability measure then. Let  $\varphi: Y \rightarrow [0, 1]$  be a  $\mathcal{X}$ -measurable function. If  $\varphi$  is continuous or it is the characteristic function of either a closed or an open subset of  $X$  then the following holds.*

(a) *If  $1 > \mu(\varphi) > m(\varphi)$  then for any  $\lambda^\mathbb{N} \ni \theta < \frac{\mu(\varphi) - m(\varphi)}{1 - m(\varphi)}$  we have*

$$\frac{D(\mu(\varphi) \parallel \theta + (1 - \theta)m(\varphi))}{-\log \theta} \leq \frac{h_m(a) - h_\mu(a)}{-\log \lambda}$$

(b) *If  $0 < \mu(\varphi) < m(\varphi)$  then for any  $\lambda^\mathbb{N} \ni \theta < \frac{m(\varphi) - \mu(\varphi)}{m(\varphi)}$  we have*

$$\frac{D(\mu(\varphi) \parallel (1 - \theta)m(\varphi))}{-\log \theta} \leq \frac{h_m(a) - h_\mu(a)}{-\log \lambda}$$

*Proof.* This follows directly from Proposition 6.1 with the constant sequence  $\mu_i = \mu$  for all  $i \in \mathbb{N}$ . The family of compact sets we take in the hypothesis of Proposition 6.1 is the collection of all compact sets for all  $\varepsilon > 0$  provided by Lemma 6.4.  $\square$

Much stronger results than the following are known in most cases, [MT94, §9], [EL10, Corollary 7.10]. Nevertheless, we think it is beneficial to state and prove this result in the general setting we are working in.

**Corollary 6.6.** *Under our standing assumptions every  $a$ -invariant and ergodic measure  $\mu$  on  $Y$  for which  $h_\nu(a) \geq h_m(a)$  has the same pushforward to the state space  $X$ .*

*Proof.* Let  $\varphi: Y \rightarrow [0, 1]$  be any  $\mathcal{X}$ -measurable continuous function with  $0 < \mu(\varphi) < 1$ . By Corollary 6.5 if  $\mu(\varphi) > \theta + (1 - \theta)m(\varphi)$  for some  $\theta \in \lambda^\mathbb{N}$  then

$$D(\mu(\varphi) \parallel \theta + (1 - \theta)m(\varphi)) \leq 0$$

which implies  $\mu(\varphi) = \theta + (1 - \theta)m(\varphi)$ . Similarly, we have  $\mu(\varphi) = (1 - \theta)m(\varphi)$  if  $\mu(\varphi) < (1 - \theta)m(\varphi)$  for some  $\theta \in \lambda^\mathbb{N}$ .

We have shown that we must have  $(1 - \theta)m(\varphi) \leq \mu(\varphi) \leq \theta + (1 - \theta)m(\varphi)$  for all  $\theta \in \lambda^\mathbb{N}$ . Taking  $\theta \rightarrow 0$  we have that  $\mu(\varphi) = m(\varphi)$  which proves the statement.  $\square$

**Theorem 6.7.** *Assume in addition to our standing assumptions that the Haar measure is a measure of maximal entropy. Let  $\mu$  be an  $a$ -invariant probability measure then. Let  $\varphi: Y \rightarrow [0, 1]$  be a continuous  $\mathcal{X}$ -measurable function, then for any  $\theta \in \lambda^\mathbb{N}$*

$$(15) \quad |\mu(\varphi) - m(\varphi)| \leq \sqrt{\frac{h_m(a) - h_\mu(a)}{-2 \log \lambda}} \sqrt{-\log \theta} + \theta$$

*Remark 6.8.* In the  $S$ -arithmetic setting, we know that the Haar measure is a measure of maximal entropy, [EL10, §7] and [MT94, §9]. If the horospherical subgroups corresponding to  $a$  generate the whole group then it actually a unique measure of maximal entropy.

*Proof.* To pass from the Kullback-Leibler divergence to the usual absolute value we use Pinsker's inequality, which in our case reduces to

$$|a - b| \leq \sqrt{\frac{1}{2} D(a \parallel b)}$$

for all  $0 \leq a, b \leq 1$ .

Assume first that  $\mu$  is ergodic and that  $0 < \mu(\varphi) < 1$ . If  $\mu(\varphi) > \theta + (1 - \theta)m(\varphi)$  we have from Corollary 6.5 and Pinsker's inequality

$$\begin{aligned} |\mu(\varphi) - m(\varphi) - \theta(1 - m(\varphi))| &\leq \sqrt{\frac{1}{2} D(\mu(\varphi) \parallel \theta + (1 - \theta)m(\varphi))} \\ &\leq \sqrt{\frac{h_m(a) - h_\mu(a)}{-2 \log \lambda}} \sqrt{-\log \theta} \end{aligned}$$

hence

$$\begin{aligned} (16) \quad |\mu(\varphi) - m(\varphi)| &\leq \sqrt{\frac{h_m(a) - h_\mu(a)}{-2 \log \lambda}} \sqrt{-\log \theta} + |\theta(1 - m(\varphi))| \\ &\leq \sqrt{\frac{h_m(a) - h_\mu(a)}{-2 \log \lambda}} \sqrt{-\log \theta} + \theta \end{aligned}$$

We can deduce in a similar manner inequality (16) also in the case that  $\mu(\varphi) < (1 - \theta)m(\varphi)$ .

The only case we have left to deal in the ergodic case is when  $(1 - \theta)m(\varphi) < \mu(\varphi) < \theta + (1 - \theta)m(\varphi)$ , but then one has the stronger inequality

$$-\theta \leq -\theta m(\varphi) \leq \mu(\varphi) - m(\varphi) \leq \theta(1 - m(\varphi)) \leq \theta$$

We have conclude the proof when  $\mu$  is ergodic,  $0 < \mu(\varphi) < 1$  and  $\theta \in \lambda^{\mathbb{N}}$  arbitrary. To remove the restriction on  $\mu(\varphi)$  look for every  $1 > \varepsilon > 0$  at  $\varphi_\varepsilon = \varepsilon + (1 - 2\varepsilon)\varphi$ . We have  $\varepsilon \leq \mu(\varphi_\varepsilon) \leq 1 - \varepsilon$ . Apply inequality (15) to  $\varphi_\varepsilon$  and take  $\varepsilon \rightarrow 0$  to see that the same holds to  $\varphi$ .

To pass to a general  $\mu$  we use the ergodic decomposition. Let  $\mathcal{E} \subseteq \mathcal{Y}$  the  $\sigma$ -algebra of  $a$ -invariant sets. The measure disintegration of  $\mu$  with respect to  $\mathcal{E}$  is an ergodic decomposition, i.e.  $\mu_y^\mathcal{E}$  is  $a$ -invariant and ergodic for  $\mu$  almost every  $y \in Y$ .

Now fix  $\theta \in \lambda^{\mathbb{N}}$  and integrate (16) over the ergodic components

$$\begin{aligned} |\mu(\varphi) - m(\varphi)| &= \left| \int_Y [\mu_y^\mathcal{E}(\varphi) - m(\varphi)] \, d\mu(y) \right| \leq \int_Y |\mu_y^\mathcal{E}(\varphi) - m(\varphi)| \, d\mu(y) \\ &\leq \int_Y \sqrt{\frac{h_m(a) - h_{\mu_y^\mathcal{E}}(a)}{-2 \log \lambda}} \, d\mu(y) \sqrt{-\log \theta} + \theta \\ &\leq \sqrt{\int_Y \frac{h_m(a) - h_{\mu_y^\mathcal{E}}(a)}{-2 \log \lambda} \, d\mu(y)} \sqrt{-\log \theta} + \theta \\ &= \sqrt{\frac{h_m(a) - h_\mu(a)}{-2 \log \lambda}} \sqrt{-\log \theta} + \theta \end{aligned}$$

Where we have used the concavity of the function  $x \mapsto \sqrt{x}$ .  $\square$

**Corollary 6.9.** *Under the hypothesis of the Theorem 6.7 there is for every  $\varepsilon > 0$  an explicitly computable constant  $C_\varepsilon > 0$  such that*

$$(17) \quad |h_m(a) - h_\mu(a)| \leq C_\varepsilon (h_m(a) - h_\mu(a))^{1/2-\varepsilon} \left[ (-\log \lambda)^{-1/2} + 2 \right]$$

Where  $C_\varepsilon$  depends only on  $\varepsilon$  and  $h_m(a)$ .

*Proof.* Denote  $D := h_m(a) - h_\mu(a)$ . If  $\sqrt{D} > \lambda$  take  $\theta = \lambda$  in Theorem 6.7

$$\begin{aligned} |\mu(\varphi) - m(\varphi)| &\leq \sqrt{\frac{D}{-2 \log \lambda}} \sqrt{-\log \theta} + \theta = \sqrt{\frac{D}{2}} + \lambda < \sqrt{D} (1/\sqrt{2} + 1) \\ &< 2\sqrt{D} \end{aligned}$$

Otherwise, let  $k \in \mathbb{N}$  such that  $\lambda^{k+1} \leq \sqrt{D} \leq \lambda^k$  and take  $\theta = \lambda^{k+1}$ . We have

$$\begin{aligned}\theta &\leq \sqrt{D} \\ -\log \theta &\leq -\frac{1}{2} \log D - \log \lambda\end{aligned}$$

and

$$\begin{aligned}(18) \quad |\mu(\varphi) - m(\varphi)| &\leq \sqrt{\frac{D \log D}{4 \log \lambda} + \frac{D}{2}} + \sqrt{D} \\ &< \sqrt{\frac{D \log D}{4 \log \lambda}} + \sqrt{D}/\sqrt{2} + \sqrt{D}\end{aligned}$$

For all  $\varepsilon > 0$  we have  $\frac{\log r}{r-\varepsilon} \rightarrow_{r \rightarrow 0+} 0$ , hence there is  $C'_\varepsilon > 0$  such that  $-\log r \leq C'_\varepsilon r^{-\varepsilon}$  for all  $0 \leq r \leq \sqrt{h_m(a)}$ .

$$|\mu(\varphi) - m(\varphi)| \leq \sqrt{C'_\varepsilon/2} D^{1/2-\varepsilon} \left[ (-\log \lambda)^{-1/2} + 2 \right]$$

The claim follows by taking  $C_\varepsilon = \max\{\sqrt{C'_\varepsilon/2}, 1\}$  and applying inequality (17) and (18).  $\square$

The following should be compared with [Rüh15, Theorem 1.1], see the discussion in §1.3.

**Theorem 6.10.** *Let  $\mathbf{G}$  be a connected simply connected absolutely almost simple linear algebraic group defined over  $\mathbb{Q}$ . Let  $S$  be a finite set of places for  $\mathbb{Q}$  including  $\infty$ , such that  $G_S := \prod_{v \in S} \mathbf{G}(\mathbb{Q}_v)$  is not compact.*

*Let  $p \in S$  be a finite place such that  $G_p := \mathbf{G}(\mathbb{Q}_p)$  is not compact. Fix  $a \in G_p \cap \mathbf{G}(\mathbb{Q})$   $\mathbb{Q}_p$ -regular and semisimple and not contained in a compact subgroup.*

*Denote by  $A < G_p$  the maximal split subtorus of a maximal torus including the element  $a$ . Let  $\Delta$  be the affine Bruhat-Tits building associated to  $G_p$ . Denote by  $H < G_p$  the pointwise stabilizer of the apartment corresponding to  $A$  in  $\Delta$ .*

*Let  $\Gamma < G_S$  be a congruence lattice and denote by  $m$  the Haar probability measure on  $\Gamma \backslash G_S$ .*

*Let  $\varphi: \Gamma \backslash G_S/H \rightarrow [0, 1]$  be a continuous function that is smooth in the  $p$ -adic coordinate, i.e. when considered as a function on  $\Gamma \backslash G_S$  it is invariant under some compact open subgroup  $K_0 < G_p$  containing  $H$ . Then for any  $\varepsilon > 0$  there is an explicitly computable constant  $C > 0$  depending only on  $\varepsilon$ ,  $G_p$ ,  $K_0$  and  $a$  such that for any  $a$ -invariant measure  $\mu$  we have*

$$|\mu(\varphi) - m(\varphi)| \leq C (h_m(a) - h_\mu(a))^{1/2-\varepsilon}$$

*Remark 6.11.* In this setting the Haar measure is a measure of maximal entropy, [MT94, §9] and [EL10, §7.8]

*Proof.* For the  $S$ -arithmetic quotients we consider there is an  $L^2$ -norm gap for the averaging operator  $T$  for *any* state space of the form  $Y/V$  where  $V < G_p$  is a compact-open subgroup. The specific result we use appears in [GMO08, §1.6]. See the discussion in §1.3 regarding bounds on matrix coefficients.

If  $\mu$  is  $a$  invariant then it is  $a^n$  invariant for all  $n \in \mathbb{N}$ . Let  $\Delta$  be the affine Bruhat-Tits building associated to  $G_p$ . Let  $\mathcal{A} \subset \Delta$  be the apartment corresponding to  $A$ . For any special vertex  $v \in \mathcal{A}$  we have associated in §4.3 an arrow subgroup  $K_{(v, a^{-2n}.v)}$  for the  $a^{-2n}$  action which stabilizes the rhomboid bounded by the corners  $v$  and  $a^{-2n}.v$ .

Fix a special vertex  $v_0 \in \mathcal{A}$  and consider the arrow subgroups  $K_n := K_{(a^n.v_0, a^{-n}.v_0)}$ . The corresponding rhomboids  $\mathcal{R}_n$  are a monotone ascending sequence of convex subsets of  $\mathcal{A}$  which cover the apartment as  $n \rightarrow \infty$ . Hence  $\bigcap_{n=1}^{\infty} K_n = H$  – the pointwise stabilizer of the apartment. By Proposition 4.8 all the state spaces  $Y/K_n$  have property (M) with respect to the right action by  $a^{2n}$ .

If  $\varphi$  is invariant under  $K_0$  then the sequence  $\{K_n \setminus K_0\}_{n=1}^{\infty}$  is a descending sequence of compact sets having empty intersection, hence there exists  $n \in \mathbb{N}$  such that  $K_n \setminus K_0 = \emptyset \Leftrightarrow K_n \subseteq K_0$  and  $\varphi$  is  $K_n$  invariant. Let  $T_{a^{2n}}$  be the averaging operator associated to  $Y/K_n$  and the right action by  $a^{2n}$ . Set  $\lambda := \|T_{a^{2n}}\|_{L_0^2(X)}$ .

We have  $\lambda < 1$  by [GMO08, §1.6]. By Corollary 6.9

$$\begin{aligned} |\mu(\varphi) - m(\varphi)| &\leq C_\varepsilon (h_m(a^{2n}) - h_\mu(a^{2n}))^{1/2-\varepsilon} \left[ (\log \lambda)^{-1/2} + 2 \right] \\ &= C_\varepsilon \sqrt{2n} \left[ (\log \lambda)^{-1/2} + 2 \right] (h_m(a) - h_\mu(a))^{1/2-\varepsilon} \end{aligned}$$

Set now  $C = C_\varepsilon \sqrt{2n} [(\log \lambda)^{-1/2} + 2]$ . □

## 7. NON-ESCAPE OF MASS

There are two flavors of related results, one where we have a sequence of measures with average mass of Bowen balls exponentially small on finer and finer scale and the second where we have a sequence of measures with high metric entropy. We shall focus only on the second case, although results of the first type can be deduced using our methods.

See §1.4 for a discussion about the relation about our theorem for non-escape of mass and previously known results in the archimedean place.

**Proposition 7.1.** *Assume in addition to our standing assumptions that the Haar measure is a measure of maximal entropy. Let  $\mu$  be an  $a$ -invariant*

probability measure. For any closed subset  $F \subset X$  we have for all  $\theta \in \lambda^{\mathbb{N}}$

$$1 - \mu(F) \leq \frac{\log \theta}{\log(\theta + (1 - \theta)(1 - m(F)))} \left[ \frac{h_m(a) - h_\mu(a)}{-\log \lambda} + \frac{1}{-\log \theta} \right]$$

*Proof.* Let  $\mathcal{E} \subseteq \mathcal{Y}$  the  $\sigma$ -algebra of  $a$ -invariant sets. The conditional measure  $\mu_y^{\mathcal{E}}$  is well-defined,  $a$ -invariant and ergodic for all  $y \in Y_0$ , with  $\mu(Y_0) = 1$ . Fix  $\theta \in \lambda^{\mathbb{N}}$ . Set  $Y_1 := \{y \in Y_0 \mid \mu_y^{\mathcal{E}}(F) < (1 - \theta)m(F)\}$ . By Corollary 6.5 for any  $y \in Y_1$

$$(19) \quad \frac{D(\mu_y^{\mathcal{E}}(F) \parallel (1 - \theta)m(F))}{-\log \theta} \leq \frac{h_m(a) - h_{\mu_y^{\mathcal{E}}}(a)}{-\log \lambda}$$

We now expand  $D(\mu_y^{\mathcal{E}}(F) \parallel (1 - \theta)m(F))$  using the explicit definition of the Kullback-Leibler divergence. For any  $a, b \in [0, 1]$

$$\begin{aligned} D(a \parallel b) &= -a \log(b) - (1 - a) \log(1 - b) - H(a) \\ &\geq -a \log(b) - (1 - a) \log(1 - b) - 1 \end{aligned}$$

where  $H(a) = -a \log(a) - (1 - a) \log(1 - a) \leq 1$  is the binary entropy function in natural base. Using this inequality and (19) we have for all  $y \in Y_1$

$$\begin{aligned} (1 - \mu_y^{\mathcal{E}}(F)) \frac{\log(\theta + (1 - \theta)(1 - m(F)))}{\log \theta} \\ \leq \mu_y^{\mathcal{E}}(F) \frac{\log((1 - \theta)m(F))}{\log \theta} + (1 - \mu_y^{\mathcal{E}}(F)) \frac{\log(\theta + (1 - \theta)(1 - m(F)))}{\log \theta} \\ \leq \frac{h_m(a) - h_{\mu_y^{\mathcal{E}}}(a)}{-\log \lambda} + \frac{1}{-\log \theta} \end{aligned}$$

Denote  $c := \frac{\log(\theta + (1 - \theta)(1 - m(F)))}{\log \theta} > 0$ . Integrating over  $Y_1$  we have

$$(20) \quad \int_{Y_1} (1 - \mu_y^{\mathcal{E}}(F)) d\mu(y) \leq c^{-1} \int_{Y_1} \frac{h_m(a) - h_{\mu_y^{\mathcal{E}}}(a)}{-\log \lambda} d\mu(y) + c^{-1} \mu(Y_1) \frac{1}{-\log \theta}$$

Set  $Y_2 := \{y \in Y_0 \mid \mu_y^{\mathcal{E}}(F) \geq (1 - \theta)m(F)\}$ , then

$$\begin{aligned} \int_{Y_2} (1 - \mu_y^{\mathcal{E}}(F)) d\mu(y) &\leq \mu(Y_2) [\theta + (1 - \theta)(1 - m(F))] \\ (21) \quad &\leq c^{-1} \int_{Y_2} \frac{h_m(a) - h_{\mu_y^{\mathcal{E}}}(a)}{-\log \lambda} + \mu(Y_2) [\theta + (1 - \theta)(1 - m(F))] \end{aligned}$$



Summing the inequalities (20) and (21) and using the linearity of the entropy function on the space of measures we have

$$\begin{aligned} 1 - \mu(F) &\leq c^{-1} \left[ \frac{h_m(a) - h_\mu(a)}{-\log \lambda} + \mu(Y_1) \frac{1}{-\log \theta} \right] \\ &\quad + \mu(Y_2) [\theta + (1 - \theta)(1 - m(F))] \\ &\leq c^{-1} \left[ \frac{h_m(a) - h_\mu(a)}{-\log \lambda} + \frac{1}{-\log \theta} \right] + \theta + (1 - \theta)(1 - m(F)) \end{aligned}$$

□

**Corollary 7.2** (Non-escape of mass). *Suppose we are given a sequence of  $a$ -invariant probability measures  $\mu_i$  on  $Y$ . Set  $h = \liminf_{i \rightarrow \infty} h_{\mu_i}(a)$ .*

*If  $\mu_i$  converges to a measure  $\mu$  in the weak-\* topology then*

$$\mu(Y) \geq 1 - \frac{h_m(a) - h}{-\log(\lambda)}$$

*Proof.* Any compact subset  $\Omega \subset Y$  is a closed subset of the one-point compactification of  $Y$ , hence  $\mu(\Omega) \geq \limsup_{i \rightarrow \infty} \mu_i(\Omega)$ . Fix a compact subset  $\Omega \subset X$ . Applying Proposition 7.1 to  $\mu_i(\Omega)$  for all  $i \in \mathbb{N}$  and taking  $i \rightarrow \infty$  we have for all  $\theta \in \lambda^{\mathbb{N}}$

$$(22) \quad 1 - \mu(\Omega) \leq \frac{\log \theta}{\log(\theta + (1 - \theta)(1 - m(\Omega)))} \left[ \frac{h_m(a) - h}{-\log \lambda} + \frac{1}{-\log \theta} \right] + \theta + (1 - \theta)(1 - m(\Omega))$$

The space  $Y$  is  $\sigma$ -compact, using the quotient map one can push forward a countable collection of compact sets covering  $Y$  to such a collection covering  $X$ , hence  $X$  is also  $\sigma$ -compact.

If we now take an ascending sequence of compact sets  $\Omega_k \subset X$  such  $\bigcup_{k=1}^{\infty} \Omega_k = X$  then taking the limit  $k \rightarrow \infty$  in inequality (22) we have <sup>20</sup>

$$1 - \mu(Y) \leq \frac{h_m(a) - h}{-\log \lambda} + \frac{1}{-\log \theta} + \theta$$

The claim follows by taking the limit  $\theta \rightarrow 0$ . □

The following definition appears here in the form of [GMO08, §2.2].

**Definition 7.3.** [Oh98]. Let  $\mathbf{G}$  be a connected absolutely almost simple group defined over  $\mathbb{Q}$ . Fix a finite rational prime  $p$  and set  $G_p := \mathbf{G}(\mathbb{Q}_p)$ . Let  $A < G_p$  be a maximal  $\mathbb{Q}_p$ -split torus,  $A^+$  a closed positive Weyl chamber.

Let  $\Phi^+$  be a system of positive roots in the set of all non-multipliable roots of  $G_p$  relative to  $A^+$ . Choose a maximal strongly orthogonal system  $\mathcal{S}$  in  $\Phi^+$  as defined and constructed in [Oh98]. Define

$$\eta = \prod_{\alpha \in \mathcal{S}} \alpha$$

---

<sup>20</sup> $\mu(X) = \mu(Y)$  and  $m(Y) = 1$

*Remark 7.4.* We refer to [Oh98, Theorem A] to a list of formulae for  $\eta$  for simple linear groups. We cite the relevant expression for  $\mathbf{G} = \mathbf{SL}_n$  from [COU01, Example 5.1]. Let  $A < \mathbf{SL}_n(\mathbb{Q}_p)$  be the diagonal subgroup and set

$$A^+ = \{\text{diag}(a_1, \dots, a_n) \mid a_i \in \mathbb{N}, a_i > 0 \text{ and } \forall 1 \leq i \leq n-1: a_{i+1} \mid a_i\}$$

then for all  $a \in A^+$

$$\eta(a) = \prod_{i=1}^{\lfloor n/2 \rfloor} \frac{|a_i|_p}{|a_{n+1-i}|_p}$$

For concreteness take  $a = \text{diag}(p^{n-1}, p^{n-3}, \dots, p^{-(n-1)})$  then

$$\begin{cases} p^{2k^2} & \text{if } n = 2k \\ p^{2k(k+1)} & \text{if } n = 2k + 1 \end{cases}$$

In contrast  $p^{h_m(a)} = p^{2n(n+1)(n+2)/3}$ .

**Theorem 7.5.** *Let  $\mathbf{G}$  be a connected simply connected absolutely almost simple linear algebraic group defined over  $\mathbb{Q}$ . Let  $S$  be a finite set of places for  $\mathbb{Q}$  including  $\infty$ , such that  $G_S := \prod_{v \in S} \mathbf{G}(\mathbb{Q}_v)$  is not compact.*

*Let  $p \in S$  be a finite place such that  $G_p := \mathbf{G}(\mathbb{Q}_p)$  is not compact, set  $r = 2$  if  $\text{rank}_{\mathbb{Q}_p} \mathbf{G} = 1$  and  $r = 1$  if the rank is higher.*

*Fix  $a \in G_p \cap \mathbf{G}(\mathbb{Q})$  a  $\mathbb{Q}_p$ -regular and semisimple element not belonging to a compact subgroup. Let  $A < G_p$  be a maximal split subtorus of a maximal torus containing  $a$ , set  $A^+$  to be the closed Weyl chamber corresponding to  $a$ . Define  $\eta$  with respect to these choices.*

*Let  $\Gamma < G_S$  be a congruence lattice and denote by  $m$  the Haar probability measure on  $\Gamma \backslash G_S$ .*

*Suppose we are given a sequence of  $a$ -invariant probability measures  $\mu_i$  on  $Y$ . Set  $h = \liminf_{i \rightarrow \infty} h_{\mu_i}(a)$ . If  $\mu_i$  converges to a measure  $\mu$  in the weak-\* topology then*

$$\mu(Y) \geq 1 - 2r \frac{h_m(a) - h}{-\log \eta(a)}$$

*Remark 7.6.* The factor of  $r = 2$  in the rank 1 case is conjectured to be redundant [GMO08, Conjecture 2.15], this is related to the Ramanujan conjecture.

*Proof.* Fix an arrow subgroup  $K_{\rightarrow} < G_p$  corresponding to the right action by  $a$  on  $Y := \Gamma \backslash G_S$ . By Proposition 4.8 the state space  $X := Y / K_{\rightarrow}$  has property (M) for the right action by  $a$ , hence for the action by  $a^n$  for all  $n \in \mathbb{N}$ . For  $n \in \mathbb{N}$  let  $T_{a^n}$  be the averaging operator induced by the  $a^n$  action.

By [GMO08, Theorem 1.19 and Lemma 2.3] for each  $\varepsilon > 0$  there exists  $c = c(\Gamma, K_{\rightarrow}, \varepsilon) > 0$  such that for all  $n \in \mathbb{N}$

$$\|T_{a^n}\|_{L_0^2(X)} \leq c\eta(a^n)^{(1/2-\varepsilon)/r} = c\eta(a)^{n(1/2-\varepsilon)/r}$$

Apply Corollary 7.2 to the action by  $a^n$  and use the fact that  $h_{\mu_i}(a^n) = nh_{\mu_i}(a)$  to see that

$$\mu(Y) \geq 1 - \frac{nh_m(a) - nh}{- [n(1/2 - \varepsilon)/r] \log \eta(a) - \log c}$$

The theorem follows by taking first the limit  $n \rightarrow \infty$  and then  $\varepsilon \rightarrow 0$ .  $\square$

## APPENDIX A. LOCAL ENTROPY

For lack of a good reference we present a proof of the following partial extension of the Brin-Katok [BK83] theorem to second countable metric spaces.

**Theorem A.1.** *Let  $(Y, d)$  be a second countable metric space with a Borel  $\sigma$ -algebra  $\mathcal{Y}$ . Let  $\mu$  be a Borel probability measure on  $Y$  and  $S: Y \rightarrow Y$  be an invertible continuous ergodic measure preserving transformation. Define for every  $y \in Y$ ,  $r > 0$  and  $n \in \mathbb{N}$*

$$B_r^{(-n, n)}(y) = \{z \in Y \mid \forall -n \leq i \leq n: d(S^i(z), S^i(y)) < r\}$$

*Then for  $\mu$ -almost every  $y \in Y$*

$$\lim_{r \rightarrow 0} \liminf_{n \rightarrow \infty} -\frac{\log \mu(B_r^{(-n, n)}(y))}{2n+1} \geq h_\mu(S)$$

*Proof.* We prove that for all  $\delta > 0$  there exists  $\varepsilon > 0$  such that  $\mu$ -almost every  $y \in Y$  for all  $n$  large enough and  $r < \varepsilon$

$$\log \mu(B_r^{(-n, n)}(y)) \geq \exp[-(2n+1)(h_\mu(S) - \delta)]$$

*Constructing a partition with  $\mu$ -continuous atoms and high entropy.* Recall that  $C \subseteq Y$  is called a  $\mu$ -continuity set if  $\mu(\partial C) = 0$ . For any point  $y \in Y$  all but countable many of the balls  $\{B_r(y) \mid r \in \mathbb{R}_{>0}\}$  are  $\mu$ -continuity sets. Hence for all  $y \in Y$  there is a countable dense subset of  $\mathbb{R}_{>0}$

$$R_y = \{r_y^1, r_y^2, \dots\}$$

such that  $B_{r_y^i}(y)$  is a  $\mu$ -continuity set for all  $i \in \mathbb{N}$ .

Let  $\{\Omega_k = B_{\rho_k}(y_0)\}_{k=1}^\infty$  be a sequence of  $\mu$ -continuous balls centered at a common point  $y_0$  with radii  $\rho_k \rightarrow_{k \rightarrow \infty} \infty$ .

Let  $\{y_j\}_{j=1}^\infty$  be a dense countable subset of  $X$ . Define the finite partition  $\xi_l$  to be composed of all the mutual refinements of the sets

$$\left\{ B_{r_{y_j}^i} \cap \Omega_k \mid 1 \leq i, j, k \leq l \right\} \cup \{Y \setminus \Omega_l\}$$

For every open ball  $B(y, r) \subset Y$  we can find a subsequence  $\{y_{j_m}\}_{m=1}^\infty$  such that  $d(y_{j_m}, y) < 1/m$ . For each  $m \in \mathbb{N}$  let  $i_m$  be such that  $|r_{y_{j_m}}^{i_m} - r| < 1/m$ . Let  $k \in \mathbb{N}$  such that  $B(y, r+2) \subset \Omega_k$ . The ball  $B_m := B_{r_{y_{j_m}}^{i_m}}(y_{j_m}) \subseteq B(y, r+2)$  belongs to  $\sigma(\xi^l)$  for  $l = \max\{i_m, j_m, k\}$ . We have

$$B(y, r) = \bigcap_{M=1}^\infty \bigcup_{m=M}^\infty B_m$$

Hence the  $\sigma$ -algebra generated by all the partitions  $\{\xi_l\}_{l=1}^\infty$  is  $\mathcal{Y}$ .

These partitions satisfy  $\sigma(\xi_l) \nearrow_{l \rightarrow \infty} \mathcal{Y}$  hence  $h_\mu(S, \xi_l) \rightarrow_{l \rightarrow \infty} h_\mu(S)$ . Moreover, all the subsets of the partitions  $\{\xi_l\}_{l=1}^\infty$  are  $\mu$ -continuity sets.

Fix  $\xi = \xi_l$  for  $l$  large enough so that

$$(23) \quad h_\mu(\xi, S) > h_\mu(S) - \delta/4$$

The proof continues from here using the same method as [BK83]; we provide the details for completeness sake.

*Picking up good atoms.* For a set  $C \subseteq Y$  and  $\varepsilon > 0$  define

$$C^{(\varepsilon)} := \{y \in Y \mid d(y, C) \leq \varepsilon\}$$

Set  $E := \bigcup_{W \in \xi} \partial W$ . Because  $\mu(E) = 0$  we have  $\mu(E^{(\varepsilon)}) \rightarrow_{\varepsilon \rightarrow 0} 0$ .

For any  $y \in Y$  let

$$I_n(y) := \left\{ -n \leq i \leq n \mid S^i y \notin E^{(\varepsilon)} \right\}$$

Define  $\zeta_n(y) := \bigcap_{i \in I_n(y)} S^{-i}(\xi)(y)$

*The relation between  $\zeta_n(y)$  and  $B_\varepsilon^{(-n, n)}(y)$ .* The set  $\zeta_n(y)$  consists of all the points whose orbit in the times  $i \in I_n(y)$  visits the same atoms of  $\xi$  as  $y$ . If  $y' \in B_\varepsilon^{(-n, n)}(y)$  then in all the times  $-n \leq i \leq n$  we have  $S^i(y') \in B_\varepsilon(S^i(y))$ . For any  $i \in I_n(y)$  we have  $B_\varepsilon(S^i(y)) \subseteq \xi(y)$ , hence  $S^i(y') \in \xi(y)$ . This proves that  $B_\varepsilon^{(-n, n)}(y) \subseteq \zeta_n(y)$ .

Thus we need only prove for  $\mu$ -almost every  $y \in Y$

$$\lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{-\log \mu(\zeta_n(y))}{2n+1} \geq h_\mu(S) - \delta$$

*Estimating the number of typical sets  $\zeta_n(y)$  containing a common point.* Define

$$\mathcal{Z}_n := \left\{ \zeta_n(y) \mid y \in Y \text{ and } |I_n(y)| > (2n+1)(1 - 2\mu(E^{(\varepsilon)})) \right\}$$

The family  $\mathcal{Z}_n$  is a collection of subsets of  $Y$ . The sets  $\zeta_n(y)$  in  $\mathcal{Z}_n$  are those for which  $y$  visits the set  $E^{(\varepsilon)}$  in the typical way. Indeed, applying the ergodic theorem to the function  $\chi_{E^{(\varepsilon)}}$  we see that for  $\mu$ -almost every  $y \in Y$  one has that  $\zeta_n(y) \in \mathcal{Z}_n$  for all large enough  $n$ .

The sets in  $\mathcal{Z}_n$  are not necessarily mutually disjoint. Fix  $y_0 \in Y$ , how many sets in  $\mathcal{Z}_n$  contain it? If  $y_0 \in \zeta_n(y)$  then for all  $i \in I_n(y)$  we have that  $\xi(S^i y) = \xi(S^i y_0)$ . Hence each such  $\zeta_n(y)$  is the intersection of at least  $(2n+1)(1 - 2\mu(E^{(\varepsilon)}))$  sets from the collection

$$\{S^{-i}(\xi)(y_0) \mid -n \leq i \leq n\}$$

This implies that the number of sets in  $\mathcal{Z}_n$  containing  $y$  is bounded above by

$$\begin{aligned} F_n(\varepsilon) &:= \sum_{m=(2n+1)(1-2\mu(E^{(\varepsilon)}))}^{2n+1} \binom{2n+1}{m} = \sum_{m=0}^{(2n+1)2\mu(E^{(\varepsilon)})} \binom{2n+1}{m} \\ &\leq \exp \left[ (2n+1)H(2\mu(E^{(\varepsilon)})) \right] \end{aligned}$$

Where  $H(a) = -a \log a - (1-a) \log(1-a)$  is the binary entropy function in natural base. What is important for us is that the bound is uniform in  $y$  and

$$(24) \quad H(2\mu(E^{(\varepsilon)})) \rightarrow_{\varepsilon \rightarrow 0} 0$$

*Estimating the number of typical sets  $\zeta_n(y)$  having atypically big measure.* We want to estimate how many sets there are in the family

$$\mathcal{Z}_n^{\text{big}} := \{C \in \mathcal{Z}_n \mid \mu(C) > \exp[-(2n+1)(h_\mu(S) - \delta)]\}$$

The sets in  $\mathcal{Z}_n^{\text{big}}$  are those having measure larger than what we expect asymptotically. Because  $\mu$ -almost every  $y \in Y$  has  $\zeta_n(y) \in \mathcal{Z}_n$  for all  $n$  large enough, we deduce that for  $\mu$ -almost every  $y \in Y$  for all  $n$  large enough if

$$(25) \quad \mu(\zeta_n(y)) > \exp[-(2n+1)(h_\mu(S) - \delta)]$$

then  $\zeta_n(y) \in \mathcal{Z}_n^{\text{big}}$ .

We now bound the size of  $\mathcal{Z}_n^{\text{big}}$ .

$$\begin{aligned} |\mathcal{Z}_n^{\text{big}}| \exp[-(2n+1)(h_\mu(S) - \delta)] &\leq \sum_{C \in \mathcal{Z}_n^{\text{big}}} \mu(C) = \int \sum_{C \in \mathcal{Z}_n^{\text{big}}} \chi_C(y) d\mu(y) \\ &\leq \int F_n(\varepsilon) d\mu(y) = F_n(\varepsilon) \end{aligned}$$

Hence

$$(26) \quad |\mathcal{Z}_n^{\text{big}}| \leq \exp \left[ (2n+1) \left( h_\mu(S) - \delta + H(2\mu(E^{(\varepsilon)})) \right) \right]$$

*Concluding the proof.* Define the auxiliary partition

$$\eta := \left\{ W \cap E^{(\varepsilon)} \mid W \in \xi \right\} \cup \left\{ Y \setminus E^{(\varepsilon)} \right\}$$

The number of atoms in  $\eta$  does not depend on  $\varepsilon$  but it has one atom,  $Y \setminus E^{(\varepsilon)}$ , whose measure goes to 1 when  $\varepsilon \rightarrow 0$ , hence

$$(27) \quad h_\mu(S, \eta) \leq h_\mu(\eta) \rightarrow_{\varepsilon \rightarrow 0} 0$$

For any partition  $\alpha$  of  $Y$  denote  $\alpha^{(-n,n)} := \bigvee_{i=-n}^n S^{-i}(\alpha)$ . By the definition of  $\zeta_n$  we see for all  $y \in Y$  that

$$\begin{aligned} (28) \quad \zeta_n(y) \cap \eta^{(-n,n)}(y) &\subseteq \xi^{(-n,n)}(y) \\ \implies \mu \left( \zeta_n(y) \cap \eta^{(-n,n)}(y) \right) &\leq \mu \left( \xi^{(-n,n)}(y) \right) \end{aligned}$$

Define

$$\begin{aligned}\mathcal{X}_n &= \left\{ \xi^{(-n,n)}(y) \mid y \in Y \text{ and } \mu(\xi^{(-n,n)}) < \exp[-(2n+1)(h_\mu(S, \xi) - \delta/4)] \right\} \\ \mathcal{H}_n &= \left\{ \eta^{(-n,n)}(y) \mid y \in Y \text{ and } \mu(\eta^{(-n,n)}) > \exp[-(2n+1)(h_\mu(S, \eta) + \delta/4)] \right\}\end{aligned}$$

These are families of typical atoms for  $\xi^{(-n,n)}$  and  $\eta^{(-n,n)}$  respectively. By the Shannon-MacMillan-Breiman for  $\mu$ -almost every  $y \in Y$  we have for all  $n$  large enough  $\xi^{(-n,n)}(y) \in \mathcal{X}_n$  and  $\eta^{(-n,n)}(y) \in \mathcal{H}_n$ . All the sets in  $\mathcal{H}_n$  are mutually disjoint, thus

$$(29) \quad |\mathcal{H}_n| \leq \exp[(2n+1)(h_\mu(S, \eta) + \delta/4)]$$

Finally let

$$\begin{aligned}\mathcal{B}_n &= \left\{ \zeta_n(y) \cap \eta^{(-n,n)}(y) \mid y \in Y, \xi^{(-n,n)}(y) \in \mathcal{X}_n \right. \\ &\quad \left. , \eta^{(-n,n)}(y) \in \mathcal{H}_n \text{ and } \zeta_n(y) \in \mathcal{Z}_n^{\text{big}} \right\}\end{aligned}$$

For any  $C \in \mathcal{B}_n$  we have by (28) and the definition of  $\mathcal{X}_n$

$$(30) \quad \mu(C) \leq \exp[-(2n+1)(h_\mu(S, \xi) - \delta/4)]$$

We can now give the main bound on the measure of the union of all the sets in  $\mathcal{B}_n$  multiplying (26), (29) and (30)

$$\begin{aligned}\mu\left(\bigcup_{C \in \mathcal{B}_n} C\right) &\leq |\mathcal{Z}_n^{\text{big}}| |\mathcal{H}_n| \exp[-(2n+1)(h_\mu(S, \xi) - \delta/4)] \\ &\leq \exp\left[-(2n+1)\left(h_\mu(S, \xi) - h_\mu(S) - h_\mu(S, \eta) + \delta/2 + H(2\mu(E^{(\varepsilon)}))\right)\right] \\ &\leq \exp\left[-(2n+1)\left(\delta/4 - h_\mu(S, \eta) + H(2\mu(E^{(\varepsilon)}))\right)\right]\end{aligned}$$

By (24) and (27) for any  $\varepsilon > 0$  small enough we have

$$|h_\mu(S, \eta) - H(2\mu(E^{(\varepsilon)}))| < \delta/8$$

and

$$\mu\left(\bigcup_{C \in \mathcal{B}_n} C\right) < \exp[-(2n+1)\delta/8]$$

□

As this bound is summable, by Borel-Cantelli  $\mu$ -almost every  $y \in Y$  belongs to no more than finitely many of these unions. Hence for  $\mu$ -almost every  $y \in Y$  one of the three conditions appearing in the definition of  $\mathcal{B}_n$  occurs only finitely many times, but that first two occur  $\mu$ -almost every  $y$  for all  $n$  large enough by the Shannon-McMillan-Breiman theorem. Hence  $\mu$ -almost every  $y \in Y$  has  $\zeta_n(y) \notin \mathcal{Z}_n^{\text{big}}$  for all  $n$  large enough, but as we have discussed before this implies by the ergodic theorem that for  $\mu$ -almost every  $y \in Y$  inequality (25) holds for all  $n$  large enough.

## APPENDIX B. HOMOGENEITY OF THE HAAR MEASURE

The following is standard, at least in the split case. We provide a proof for completeness sake.

**Proposition B.1.** *Let  $G = G_0 \times G_p$  be a product of locally compact groups and  $\Gamma < G$  a lattice. Suppose that  $G_p = \mathbf{G}(\mathbb{Q}_p)$  for  $\mathbf{G}$  a linear reductive group defined over  $\mathbb{Q}$  and  $p$  a finite rational prime.*

*Let  $a \in G_p$  be semisimple. By abuse of notation denote by  $m$  the Haar measure on  $G$  and  $\Gamma \backslash G$  and normalize  $m$  on  $G$  so it induces a probability measure on  $\Gamma \backslash G$ . Then there is a base for the topology of  $G$  around the identity consisting of  $(a, h_m(a))$ -homogeneous neighborhoods.*

*Proof.* It suffices to produce an  $(a, h_m(a))$ -homogeneous base for the topology of  $G_p$ . In order to conjugate by  $a$  instead of  $a^{-1}$  we prove the theorem for  $a^{-1}$ -Bowen balls. Evidently,  $B_{a^{-1}}^{(s,t)} = B_a^{(-t,-s)}$  and  $h_m(a^{-1}) = h_m(a)$ .

We consider  $\mathbf{G}$  as defined over  $\mathbb{Q}_p$ , i.e. replace  $\mathbf{G}$  with its extension of scalars to  $\mathbb{Q}_p$ . Let  $\mathbf{A} < \mathbf{G}$  be a maximal torus whose  $\mathbb{Q}_p$  points include  $a$  and denote by  $\mathbb{L}$  the splitting field of  $\mathbf{A}$ , it is a finite Galois field extension of  $\mathbb{Q}_p$ . Let  $\mathfrak{G} := \text{Gal}(\mathbb{L}/\mathbb{Q}_p)$  and denote by  $e$  the ramification index of  $\mathbb{L}/\mathbb{Q}_p$ . Let  $w$  be the single place of  $\mathbb{L}$  extending  $p$ .

Let  $\mathfrak{g}$  be the Lie algebra  $\mathbf{G}$  and denote  $\mathfrak{g}_p := \mathfrak{g}(\mathbb{Q}_p)$  and  $\mathfrak{g}_w := \mathfrak{g}(\mathbb{L})$ . The  $\mathbb{L}$ -vector space  $\mathfrak{g}_w$  is equipped with a Galois structure for  $\mathfrak{G}$  such that  $\mathbb{Q}_p$ -vector space  $\mathfrak{g}_p$  is the set of fixed points for the action of  $\mathfrak{G}$ .

Consider the decomposition of  $\mathfrak{g}_w$  into  $\text{Ad}_a$ -eigenspaces with eigenvalue  $\lambda$

$$\mathfrak{g}_w = \bigoplus_{\lambda} \mathfrak{g}_{w,\lambda}$$

Let  $[\lambda]$  be the Galois orbit of  $\lambda$ . Denote the sum of all eigenspaces with eigenvalue Galois conjugate to  $\lambda$  by

$$\mathfrak{g}_{w,[\lambda]} = \bigoplus_{\sigma \in \mathfrak{G}} \mathfrak{g}_{w,\sigma(\lambda)}$$

The space  $\mathfrak{g}_{w,[\lambda]} < \mathfrak{g}_w$  is  $\mathfrak{G}$ -invariant hence it defined over  $\mathbb{Q}_p$ , let  $\mathfrak{g}_{[\lambda]} < \mathfrak{g}_p$  be the subspace of  $\mathfrak{G}$  fixed vectors, then  $\mathfrak{g}_{[\lambda]} \otimes_{\mathbb{Q}_p} \mathbb{L} = \mathfrak{g}_{w,[\lambda]}$ .

Let  $X_{1,[\lambda]}, \dots, X_{k,[\lambda]}$  be a  $\mathfrak{G}$ -stable eigenbase for  $\mathfrak{g}_{w,[\lambda]}$ . We construct an  $\mathcal{O}_{\mathbb{L}}$  lattice in  $\mathfrak{g}_{w,[\lambda]}$

$$L_{w,[\lambda]} = \bigoplus_{i=1}^k X_{i,[\lambda]} \mathcal{O}_{\mathbb{L}}$$

If  $e$  is the ramification index then there exists  $F_{[\lambda]} \in \mathbb{Q}_p$  such that  $\lambda^e = F_{[\lambda]} \omega$  for some  $\omega \in \mathcal{O}_{\mathbb{L}}^{\times}$ . Any  $\sigma \in \mathfrak{G}$  acts by  $\sigma(\lambda^e) = F_{[\lambda]} \sigma(\omega)$  with  $\sigma(\omega) \in \mathcal{O}_{\mathbb{L}}^{\times}$ . Thus for all  $n \in \mathbb{Z}$

$$\text{Ad}_{a^{en}}(L_{w,[\lambda]}) = F_{[\lambda]}^n L_{w,[\lambda]}$$

The lattice  $L_{w,[\lambda]}$  is invariant under  $\mathfrak{G}$ , let  $L_{[\lambda]}$  be the set of  $\mathfrak{G}$  fixed points in  $L_{w,[\lambda]}$ . By definition  $L_{[\lambda]} = \mathfrak{g}_{[\lambda]} \cap L_{w,[\lambda]}$ . This is an intersection of a compact open subset of  $\mathfrak{g}_{w,[\lambda]}$  with a subspace, hence  $L_{[\lambda]}$  is a compact open subset of  $\mathfrak{g}_{[\lambda]}$ . Moreover,  $L_{[\lambda]}$  is clearly a  $\mathbb{Z}_p$  submodule of the finitely

generated  $\mathbb{Z}_p$  module  $L_{w,[\lambda]}$ . Because  $\mathbb{Z}_p$  is Noetherian we conclude that  $L_{[\lambda]}$  is a lattice of full rank in  $\mathfrak{g}_{[\lambda]}$ .

We now show for all  $n \in \mathbb{Z}$  that  $\text{Ad}_{a^{en}}(L_{[\lambda]}) = F_{[\lambda]}^n L_{[\lambda]}$ . Fix  $n \in \mathbb{Z}$ . The two  $\mathbb{L}$ -linear automorphisms  $\text{Ad}_{a^{en}}$  and  $F_{[\lambda]}^n \text{Id}$  of  $\mathfrak{g}_{w,[\lambda]}$  are defined over  $\mathbb{Q}_p$ , hence  $\text{Ad}_{a^{en}}(L_{[\lambda]})$  and  $F_{[\lambda]}^n L_{[\lambda]}$  are the sets of Galois fixed points in  $\text{Ad}_{a^{en}}(L_{w,[\lambda]}) = F_{[\lambda]}^n L_{w,[\lambda]}$ . This proves  $\text{Ad}_{a^{en}}(L_{[\lambda]}) = F_{[\lambda]}^n L_{[\lambda]}$ .

Let  $|\cdot|_w$  be an absolute value corresponding to  $w$  on  $\mathbb{L}$ . Galois conjugate eigenvalues have the same absolute value. If  $|\lambda|_w \geq 1$  then for any  $t_2 \geq t_1$  we have  $\text{Ad}_{a^{t_1}} L_{w,[\lambda]} \subseteq \text{Ad}_{a^{t_2}} L_{w,[\lambda]}$ , by considering  $\mathfrak{G}$  fixed points this implies also  $\text{Ad}_{a^{t_1}} L_{[\lambda]} \subseteq \text{Ad}_{a^{t_2}} L_{[\lambda]}$ . Similarly, if  $|\lambda|_w \leq 1$  then for all  $s_2 \leq s_1$  we have  $\text{Ad}_{a^{s_1}} L_{[\lambda]} \subseteq \text{Ad}_{a^{s_2}} L_{[\lambda]}$ . If  $|\lambda|_w = 1$  then  $L_{[\lambda]}$  is invariant under  $\text{Ad}_a$ .

Split the eigenvalue classes  $[\lambda]$  into three groups

$$\begin{aligned}\Lambda_+ &:= \{[\lambda] \mid |\lambda|_w > 1\} \\ \Lambda_0 &:= \{[\lambda] \mid |\lambda|_w = 1\} \\ \Lambda_- &:= \{[\lambda] \mid |\lambda|_w < 1\}\end{aligned}$$

Define accordingly

$$\begin{aligned}\mathfrak{L}_+ &:= \bigoplus_{[\lambda] \in \Lambda_+} L_{[\lambda]} \\ \mathfrak{L}_0 &:= \bigoplus_{[\lambda] \in \Lambda_0} L_{[\lambda]} \\ \mathfrak{L}_- &:= \bigoplus_{[\lambda] \in \Lambda_-} L_{[\lambda]}\end{aligned}$$

Set also  $\mathfrak{g}_+$ ,  $\mathfrak{g}_0$  and  $\mathfrak{g}_-$  to be the corresponding subspaces of  $\mathfrak{g}_p$ .

Fix  $\mathfrak{U} \subset \mathfrak{g}_p$  a neighborhood of 0 on which  $\exp$  is an isometry. Let  $l \in \mathbb{N}$  be large enough so that  $p^l L_{[\lambda]} \subseteq \mathfrak{U}$  for all  $[\lambda]$ . Define the zero neighborhood  $\mathfrak{B}_l := p^l \mathfrak{L}_+ \oplus p^l \mathfrak{L}_0 \oplus p^l \mathfrak{L}_- \subseteq \mathfrak{U}$  and set  $B_l := \exp \mathfrak{B}_l$ . The sets  $B_l$  for all  $l$  large enough form a base for the topology around the identity and we now prove that they possess the required homogeneity property.

Using the inclusion relation between  $a$ -conjugates of lattices  $L_{[\lambda]}$  and the action of  $\text{Ad}_{a^{en}}$  on these lattices we deduce for  $s \leq es' \leq 0 \leq et' \leq t$

$$\mathfrak{B}_l^{(s,t)} := \bigcap_{s \leq n \leq t} \text{Ad}_{a^n}(\mathfrak{B}_l) \supseteq p^l \mathfrak{L}_0 \oplus p^l \bigoplus_{[\lambda] \in \Lambda^+} F_{[\lambda]}^{t'} L_{[\lambda]} \oplus p^l \bigoplus_{[\lambda] \in \Lambda^-} F_{[\lambda]}^{s'} L_{[\lambda]}$$

hence

$$\begin{aligned}\text{vol}(\mathfrak{B}_l^{(s,t)}) &\geq \text{vol}(\mathfrak{B}_l) \left( \prod_{[\lambda] \in \Lambda^+} |F_{[\lambda]}|_p \right)^{s'} \left( \prod_{[\lambda] \in \Lambda^-} |F_{[\lambda]}|_p \right)^{t'} \\ &= \text{vol}(\mathfrak{B}_l) |\det \text{Ad}_{a^e}|_{\mathfrak{g}_+}^{s'} |\det \text{Ad}_{a^e}|_{\mathfrak{g}_-}^{t'}\end{aligned}$$

The determinants of the restriction of  $\text{Ad}_a$  to  $\mathfrak{g}_+$  and  $\mathfrak{g}_-$  are reciprocal because each character of  $\mathbf{A}$  that is a root has its negative as a root as



well. Fixing  $s \leq 0 \leq t$  and taking  $s', t'$  extremal we have  $\text{vol}(\mathfrak{B}_l^{(s,t)}) \geq \text{vol}(\mathfrak{B}_l) \exp[(t - s + 2e) \log \det | \text{Ad}_a \upharpoonright_{\mathfrak{g}_-} |_p]$ . Because  $\exp$  is an isometry on  $\mathfrak{U}$  this implies a similar volume estimate for  $B_l^{(s,t)}$ .

By [EL10, §7.9] one has  $h_m(a) = -\log \det | \text{Ad}_a \upharpoonright_{\mathfrak{g}_-} |_p$ .  $\square$

*Remark B.2.* It is easy to deduce from the proof a lower bound on  $B_l^{(s,t)}$  with the same exponent  $-\log \det | \text{Ad}_a \upharpoonright_{\mathfrak{g}_-} |_p$ . Using Theorem A.1 and [ELMV09, Proposition 3.2] that would imply  $h_m(a) = -\log \det | \text{Ad}_a \upharpoonright_{\mathfrak{g}_-} |_p$  without using leafwise measure as in [EL10].

## REFERENCES

- [AFWZ95] Noga Alon, Uriel Feige, Avi Wigderson, and David Zuckerman. Derandomized graph products. *Comput. Complexity*, 5(1):60–75, 1995.
- [BK83] M. Brin and A. Katok. On local entropy. In *Geometric dynamics (Rio de Janeiro, 1981)*, volume 1007 of *Lecture Notes in Math.*, pages 30–38. Springer, Berlin, 1983.
- [BS91] M. Burger and P. Sarnak. Ramanujan duals. II. *Invent. Math.*, 106(1):1–11, 1991.
- [BT84] F. Bruhat and J. Tits. Groupes réductifs sur un corps local. II. Schémas en groupes. Existence d’une donnée radicielle valuée. *Inst. Hautes Études Sci. Publ. Math.*, (60):197–376, 1984.
- [Che11] Yitwah Cheung. Hausdorff dimension of the set of singular pairs. *Ann. of Math. (2)*, 173(1):127–167, 2011.
- [Chi95] Patrick Chiu. Covering with Hecke points. *J. Number Theory*, 53(1):25–44, 1995.
- [CLLM12] Kai-Min Chung, Henry Lam, Zhenming Liu, and Michael Mitzenmacher. Chernoff-Hoeffding bounds for Markov chains: generalized and simplified. In *29th International Symposium on Theoretical Aspects of Computer Science*, volume 14 of *LIPICs. Leibniz Int. Proc. Inform.*, pages 124–135. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2012.
- [Clo03] Laurent Clozel. Démonstration de la conjecture  $\tau$ . *Invent. Math.*, 151(2):297–328, 2003.
- [COU01] Laurent Clozel, Hee Oh, and Emmanuel Ullmo. Hecke operators and equidistribution of Hecke points. *Invent. Math.*, 144(2):327–351, 2001.
- [CU04] Laurent Clozel and Emmanuel Ullmo. Équidistribution des points de Hecke. In *Contributions to automorphic forms, geometry, and number theory*, pages 193–254. Johns Hopkins Univ. Press, Baltimore, MD, 2004.
- [DZ10] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2010. Corrected reprint of the second (1998) edition.
- [EK12] Manfred Einsiedler and Shirali Kadyrov. Entropy and escape of mass for  $\text{SL}_3(\mathbb{Z}) \backslash \text{SL}_3(\mathbb{R})$ . *Israel J. Math.*, 190:253–288, 2012.
- [EKL06] Manfred Einsiedler, Anatole Katok, and Elon Lindenstrauss. Invariant measures and the set of exceptions to Littlewood’s conjecture. *Ann. of Math. (2)*, 164(2):513–560, 2006.
- [EKP11] Manfred Einsiedler, Shirali Kadyrov, and Anke Pohl. Escape of mass and entropy for diagonal flows in real rank one situations. 2011. preprint. <http://arxiv.org/abs/1110.0910>.

- [EL10] M. Einsiedler and E. Lindenstrauss. Diagonal actions on locally homogeneous spaces. In *Homogeneous flows, moduli spaces and arithmetic*, volume 10 of *Clay Math. Proc.*, pages 155–241. Amer. Math. Soc., Providence, RI, 2010.
- [EL15] Manfred Einsiedler and Elon Lindenstrauss. On measures invariant under tori on quotients of semisimple groups. *Ann. of Math. (2)*, 181(3):993–1031, 2015.
- [ELMV09] Manfred Einsiedler, Elon Lindenstrauss, Philippe Michel, and Akshay Venkatesh. Distribution of periodic torus orbits on homogeneous spaces. *Duke Math. J.*, 148(1):119–174, 2009.
- [ELMV12] Manfred Einsiedler, Elon Lindenstrauss, Philippe Michel, and Akshay Venkatesh. The distribution of closed geodesics on the modular surface, and Duke’s theorem. *Enseign. Math. (2)*, 58(3-4):249–313, 2012.
- [EMV13] Jordan S. Ellenberg, Philippe Michel, and Akshay Venkatesh. Linnik’s ergodic method and the distribution of integer points on spheres. In *Automorphic representations and L-functions*, volume 22 of *Tata Inst. Fundam. Res. Stud. Math.*, pages 119–185. Tata Inst. Fund. Res., Mumbai, 2013.
- [Gar97] Paul Garrett. *Buildings and classical groups*. Chapman & Hall, London, 1997.
- [GI63] O. Goldman and N. Iwahori. The space of  $\mathfrak{p}$ -adic norms. *Acta Math.*, 109:137–177, 1963.
- [Gil98] David Gillman. A Chernoff bound for random walks on expander graphs. *SIAM J. Comput.*, 27(4):1203–1220, 1998.
- [GJ78] Stephen Gelbart and Hervé Jacquet. A relation between automorphic representations of  $GL(2)$  and  $GL(3)$ . *Ann. Sci. École Norm. Sup. (4)*, 11(4):471–542, 1978.
- [GMO08] Alex Gorodnik, François Maucourant, and Hee Oh. Manin’s and Peyre’s conjectures on rational points and adelic mixing. *Ann. Sci. Éc. Norm. Supér. (4)*, 41(3):383–435, 2008.
- [HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561 (electronic), 2006.
- [Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.
- [IM65] N. Iwahori and H. Matsumoto. On some Bruhat decomposition and the structure of the Hecke rings of  $\mathfrak{p}$ -adic Chevalley groups. *Inst. Hautes Études Sci. Publ. Math.*, (25):5–48, 1965.
- [JL70] H. Jacquet and R. P. Langlands. *Automorphic forms on  $GL(2)$* . Lecture Notes in Mathematics, Vol. 114. Springer-Verlag, Berlin-New York, 1970.
- [Kad12a] Shirali Kadyrov. Entropy and escape of mass for Hilbert modular spaces. *J. Lie Theory*, 22(3):701–722, 2012.
- [Kad12b] Shirali Kadyrov. Positive entropy invariant measures on the space of lattices with escape of mass. *Ergodic Theory Dynam. Systems*, 32(1):141–157, 2012.
- [Kad14] Shirali Kadyrov. Effective uniqueness of parry measure and exceptional sets in ergodic theory, 2014. preprint. <http://arxiv.org/abs/1409.0946>.
- [Kah91] Nabil Kahale. Better expansion for Ramanujan graphs. In *32nd Annual Symposium on Foundations of Computer Science (San Juan, PR, 1991)*, pages 398–404. IEEE Comput. Soc. Press, Los Alamitos, CA, 1991.
- [Kah97] Nabil Kahale. Large deviation bounds for Markov chains. *Combin. Probab. Comput.*, 6(4):465–474, 1997.
- [KKLM14] S. Kadyrov, D. Y. Kleinbock, E. Lindenstrauss, and G. A. Margulis. Singular systems of linear forms and non-escape of mass in the space of lattices. 2014. preprint. <http://arxiv.org/abs/1407.5310>.
- [KP15] S. Kadyrov and A. Pohl. Amount of failure of upper-semicontinuity of entropy in non-compact rank-one situations, and hausdorff dimension. *Ergodic Theory and Dynamical Systems*, FirstView:1–25, 10 2015.

- [KR48] M. G. Kreĭn and M. A. Rutman. Linear operators leaving invariant a cone in a Banach space. *Uspehi Matem. Nauk (N. S.)*, 3(1(23)):3–95, 1948.
- [KR50] M. G. Kreĭn and M. A. Rutman. Linear operators leaving invariant a cone in a Banach space. *Amer. Math. Soc. Translation*, 1950(26):128, 1950.
- [Lez98] Pascal Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8(3):849–867, 1998.
- [Lin68] Yu. V. Linnik. *Ergodic properties of algebraic fields*. Translated from the Russian by M. S. Keane. *Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 45*. Springer-Verlag New York Inc., New York, 1968.
- [Lin06] Elon Lindenstrauss. Invariant measures and arithmetic quantum unique ergodicity. *Ann. of Math. (2)*, 163(1):165–219, 2006.
- [LP04] Carlos A. León and François Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *Ann. Appl. Probab.*, 14(2):958–970, 2004.
- [Mar04] Grigoriy A. Margulis. *On some aspects of the theory of Anosov systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2004. With a survey by Richard Sharp: Periodic orbits of hyperbolic flows, Translated from the Russian by Valentina Vladimirovna Szulikowska.
- [MT94] G. A. Margulis and G. M. Tomanov. Invariant measures for actions of unipotent groups over local fields on homogeneous spaces. *Invent. Math.*, 116(1-3):347–392, 1994.
- [Oh98] Hee Oh. Tempered subgroups and representations with minimal decay of matrix coefficients. *Bull. Soc. Math. France*, 126(3):355–380, 1998.
- [Oh02] Hee Oh. Uniform pointwise bounds for matrix coefficients of unitary representations and applications to Kazhdan constants. *Duke Math. J.*, 113(1):133–192, 2002.
- [Pol11] Fabrizio Polo. *Equidistribution in chaotic dynamical systems*. PhD thesis, The Ohio State University, 2011. PhD Thesis.
- [Rog90] Jonathan D. Rogawski. *Automorphic representations of unitary groups in three variables*, volume 123 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 1990.
- [Rüh15] Rene Rühr. Effectivity of uniqueness of the maximal entropy measure on  $p$ -adic homogeneous spaces. *Ergodic Theory and Dynamical Systems*, FirstView:1–17, 10 2015.
- [Sar91] Peter C. Sarnak. Diophantine problems and linear groups. In *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990)*, pages 459–471. Math. Soc. Japan, Tokyo, 1991.
- [Str74] Raimond A. Struble. Metrics in locally compact groups. *Compositio Math.*, 28:217–222, 1974.
- [Wag08] Roy Wagner. Tail estimates for sums of variables sampled by a random walk. *Combin. Probab. Comput.*, 17(2):307–316, 2008.

THE EINSTEIN INSTITUTE OF MATHEMATICS, EDMOND J. SAFRA CAMPUS, THE HEBREW UNIVERSITY OF JERUSALEM, GIVAT RAM, JERUSALEM, 9190401, ISRAEL